PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N | A2

(11) International Publication Number: **WO 00/53723**

(43) International Publication Date: 14 September 2000 (14.09.00)

(21) International Application Number: PCT/US00/05875

(22) International Filing Date: 7 March 2000 (07.03.00)

(30) Priority Data:
09/265,585 10 March 1999 (10.03.99) US

(71) Applicant: NEW YORK UNIVERSITY [US/US]; 550 First Avenue, New York, NY 10016 (US).

(72) Inventors: BENFEY, Philip, N.; 3 Washington Square Village #6A, New York, NY 10012–1803 (US). DI LAURENZIO, Laura; 1737 York Avenue #2G, New York, NY 10128 (US). WYSOCKA-DILLER, Joanna; 710 West End Avenue #5–A, New York, NY 10027 (US). MALAMY, Jocelyn, E.; 14 Washington Place #4–D, New York, NY 10003 (US). PYSH, Leonard; 14 Washington Place #2–B, New York, NY 10003 (US). HELARIUTTA, Yrjo; 14 Washington Place #8–L, New York, NY 10003 (US). BRUCE, Wesley; 4625 96th Street, Urbandale, IA 50322 (US). LIM, Jun; 4 Washington Square Village, Apt. #9K, New York, NY 10012 (US).

(74) Agents: REIN, Barry, D. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

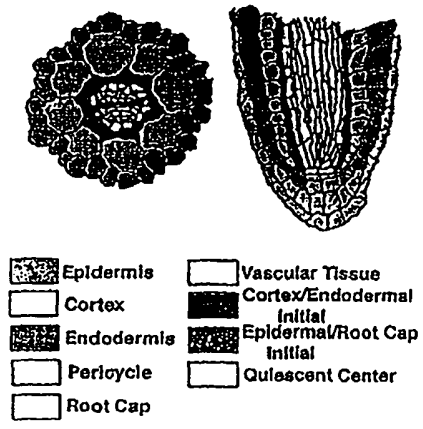
(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

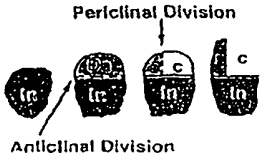
(54) Title: *SCARECROW* GENE, PROMOTER AND USES THEREOF

(57) Abstract

The structure and function of a regulatory gene, *SCARECROW* (*SCR*), is described. The *SCR* gene is expressed specifically in root progenitor tissues of embryos, and in roots and stems of seedlings and plants. *SCR* expression controls cell division of certain cell types in roots and affects the organization of root and stem tissues, and affects gravitropism of aerial structures. The invention relates to the *SCR* gene, *SCR*–like genes, *SCR* gene products, (including but not limited to transcriptional products such as mRNAs, antisense, and ribozyme molecules, and translational products such as the *SCR* protein, polypeptides, peptides and fusion proteins related thereto), antibodies to *SCR* gene products, *SCR* promoters and regulatory regions and the use of the foregoing to improve agronomically valuable plants.



A



B

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

SCARECROW GENE, PROMOTER AND USES THEREOF

This application is a continuation-in-part of co-pending Application No. 08/842,445, filed April 24, 1997, which is a continuation-in-part of Application No. 08/638,617, filed April 26, 1996, now abandoned, the disclosures of which are herein incorporated by reference in their entirety.

This invention was made with government support under grant number: GM43778 awarded by the National Institute of Health. The government may have certain rights in the invention.

1. INTRODUCTION

The present invention generally relates to the *SCARECROW* (*SCR*) gene family and their promoters. The invention more particularly relates to ectopic expression of members of the *SCARECROW* gene family in transgenic plants to artificially modify plant structures. The invention also relates to utilization of the *SCARECROW* promoter for tissue and organ specific expression of heterologous gene products.

2. BACKGROUND OF THE INVENTION

Asymmetric cell divisions, in which a cell divides to give two daughters with different fates, play an important role in the development of all multicellular organisms. In plants, because there is no cell migration, the regulation of asymmetric cell divisions is of heightened importance in determining organ morphology. In contrast to animal embryogenesis, most plant organs are not formed during embryogenesis. Rather, cells that form the apical meristems are set aside at the shoot and root poles. These reservoirs of stem cells are considered to be the source of all post-embryonic organ development in plants. A fundamental question in developmental biology is how meristems function to generate plant organs.

2.1. ROOT DEVELOPMENT

Root organization is established during embryogenesis. This organization is propagated during postembryonic development by the root meristem. Following germination, the development of the postembryonic root is a continuous process, wherein a series of initials or stem cells continuously divide to perpetuate the pattern established in the embryonic root (Steeves & Sussex, 1972, Patterns in Plant Development, Englewood Cliffs, NJ: Prentice-Hall, Inc.).

2.1.1. ARABIDOPSIS ROOT DEVELOPMENT

Due to the organization of the Arabidopsis root, it is possible to follow the fate of cells from the meristem to maturity and identify the progenitors of each cell type (Dolan et al., 1993, Development 119:71-84). The Arabidopsis root is a relatively simple and well characterized organ. The radial organization of the mature tissues in the Arabidopsis root has been likened to tree rings with the epidermis, cortex, endodermis and pericycle forming radially symmetric cell layers that surround the vascular cylinder (FIG. 1A). See also Dolan et al., 1993, Development 119:71-84. These mature tissues are derived from four sets of stem cells or initials: i) the columella root cap initial; ii) the pericycle/vascular initial; iii) the epidermal/lateral root cap initial; and iv) the cortex/endodermal initial (Dolan et al., 1993, Development 119:71-84). It has been shown that these initials undergo asymmetric divisions (Scheres et al., 1995, Development 121:53-62). The cortex/endodermal initial, for example, first divides anticlinally (in a transverse orientation) (FIG. 1B). This asymmetric division produces another initial and a daughter cell. The daughter cell, in turn, expands and then divides periclinally (in the longitudinal orientation) (FIG. 1B). This second asymmetric division produces the progenitors of the endodermis and the cortex cell lineages (FIG. 1B).

Furthermore, root radial organization in Arabidopsis is produced by three distinct developmental strategies. First, primary roots employ stem cells, wherein initials undergo asymmetric divisions first to regenerate themselves and then to generate the cell lineages of the root (Fig.1B). Second, in the embryo, sequential asymmetric divisions subdivide pre-existing tissue to form the cell layers of the embryonic root. Finally, lateral roots are formed by a strategy of cell proliferation that originates in differentiated tissues. Remarkably, within a given species, all three strategies result in roots with a nearly identical radial organization.

2.1.2. MAIZE ROOT DEVELOPMENT

The root organization of *Zea mays* (maize), which is a very well characterized member of the grass family, is far more complex than the root organization in Arabidopsis. The root system of maize consists of primary, embryonic, lateral, seminal lateral and adventitious roots. Both primary and seminal lateral roots are formed during embryogenesis, wherein the primary root is the first root to emerge during germination, followed by the seminal lateral roots formed at the scutellar nodal region (Freeling, M. and Walbot, V. (1994), The Maize Handbook, (New York: Springer-Verlag); Hetz, W. et al., (1996), Plant J. 10:845-857). Both crown and prop roots which develop post-embryonically are shoot-borne roots, often termed "adventitious". However, since these roots are part of the normal development of the plant, they are not, strictly speaking, adventitious roots, which are typically formed as a result of injury or hormone treatment. Crown roots, representing the major roots of the mature plant, are formed at consecutive early nodes of the stem beginning with the coleoptilar nodes. Later in development, brace or prop roots emerge from nodes above the soil level (Freeling, M. and Walbot, V. (1994), The Maize Handbook, (New York: Springer-Verlag); Hetz, W. et al., (1996), Plant J. 10:845-857).

Currently, there are two notably different types of organization of root apical meristems: an open and a closed meristem. In an "open" meristem, the cell files of the mature tissues cannot be traced with much confidence to distinct initials, and the incipient tissues do not appear to have discrete boundary walls between the root proper and the root cap (Clowes, F. A. L., 1981, Ann. Bot. 48:761-767). Therefore, the interpretation of the organization of the open meristem has been problematic (Clowes, F. A. L., 1981, Ann. Bot. 48:761-767). In a "closed" meristem, however, since files of cells converge onto a pole at the root apex, it is easy to identify discrete layers in median longitudinal sections (Clowes, F. A. L., 1981, Ann. Bot. 48:761-767).

Both Arabidopsis and maize roots show characteristics of the closed meristem (FIG. 23). However, there are important differences. In maize roots, the root apical meristem consists of three independent layers of initials. One gives rise to the stele, the second gives rise to epidermis, cortex and endodermis and the third generates the root cap, whereas in the Arabidopsis root apical meristem, the epidermis shares a common initial with the lateral root cap (Esau, K., 1977, Anatomy of Seed Plants. 2nd ed. (New York: John Wiley & Sons); Esau, K., 1953, Plant Anatomy. (New York: John Wiley & Sons)).

Primary organization of the root apical meristem in maize occurs during embryogenesis, (Steeves, T. A. and Sussex, I. M., (1989), Patterns in plant development., 2nd ed. (Cambridge University Press)) as in Arabidopsis. There are three main phases in embryo development in maize (FIG. 24) (Freeling, M. and Walbot, V. (1994), The Maize Handbook, (New York: Springer-Verlag); Steeves, T. A. and Sussex, I. M., (1989), Patterns in plant development., 2nd ed., (Cambridge University Press); Sheridan, W. F. and Clark, J. K., (1993), Plant J. 3:347-358). As in Arabidopsis, the very first division of the zygote establishes the initial asymmetry of the embryo (FIG. 24A). However, unlike Arabidopsis, embryonic development in maize is characterized

by rather irregular cell divisions (Sheridan, W. F. and Clark, J. K., (1993), Plant J. 3:347-358). During the first phase, the apical-basal asymmetry of the embryo is established, and the embryo is regionalized into suspensor and embryo proper (FIG. 24B-C). During the second phase, radial asymmetry appears and the embryonic axis and meristems are established (FIG. 24D-E) (Clowes, F. A. L., (1978), New Phytol. 80:409-419). Finally, during the third phase, vegetative structures such as embryonic roots and leaves are elaborated (FIG. 24F-G) (Sheridan, W. F. and Clark, J. K., (1993), Plant J. 3:347-358).

2.1.3. THE QUIESCENT CENTER

The quiescent center (QC) of root apical meristems of angiosperms is a population of mitotically inactive cells. In the QC of the primary root of maize, for example, the average duration of a mitotic cycle is about 200 hours compared with only 12 hours in the cells just below the QC and 28 hours in the cells just above the QC (Clowes, F. A. L., (1961), J. Exp. Bot. 9:229-238). Moreover, there are also reductions in the rates of synthesis of DNA and protein, and corresponding reductions in the amounts of DNA and RNA per cell (Clowes, F. A. L., (1956), New Phytol. 55:29-34).

Although the precise role of the QC has remained speculative, it is generally accepted that cells within the QC are undifferentiated and, other than the anatomical pattern of cell files, lacking in radial pattern information. This theory has been supported by ablation studies performed in Arabidopsis, wherein, complete laser ablation of the four central cells in the Arabidopsis QC led to subsequent restoration of the QC by cells of the stele. Furthermore, laser ablation of only one or two cells in the QC resulted in differentiation of surrounding initial cells. Analysis of the *hobbit* mutants further supports these observations. In the *hobbit* mutants, there is no functional QC, leading all cortex initials to divide into cortex and endodermis during embryogenesis (van den Berg, C., et al., (1995), Nature

378:62-65). Taken together, it is suggested that the QC suppresses differentiation of surrounding initials in the range of a single cell (van den Berg, C., et al., (1995), Nature 378:62-65).

In maize, on which the contemporary view of the role of the QC is based (Feldman, L. J., (1984), Amer. J. Bot. 71:1308-1314; Freeling, M. and Walbot, V., (1994), The maize handbook.(New York: Springer-Verlag)), surgical and tissue culture systems were developed to study the organization process of root apical meristems (Feldman, L. J., (1976), Planta 128:207-212). Following removal of the QC, the remaining root regenerates a new root tip. This process appears to involve *de novo* organization of the QC and the apical meristem (Feldman, L. J., (1976), Planta 128:207-212). In addition, the excised QC itself is capable of generating a new root (Feldman, L. J. and Torrey, J. G., (1976), Amer. J. Bot. 63:345-355). This suggests that there is indeed sufficient radial pattern information in the QC to allow the regeneration of more or less intact roots.

2.2. GENES REGULATING ROOT STRUCTURE

Mutations that disrupt the asymmetric divisions of the cortex/endodermal initial have been identified and characterized (Benfey et al., 1993, Development 119:57-70; Scheres et al., 1995, Development 121:53-62). *short-root* (*shr*) and *scarecrow* (*scr*) mutants are missing a cell layer between the epidermis and the pericycle. In both types of mutants, the cortex/endodermal initial divides anticlinally, but the subsequent periclinal division that increases the number of cell layers does not take place (Benfey et al., 1993, Development 119:57-70; Scheres et al., 1995, Development 121:53-62). The defect is first apparent in the embryo and it extends throughout the entire embryonic axis, which includes the embryonic root and hypocotyl (Scheres et al., 1995, Development 121:53-62). This is true also for other radial organization mutants characterized to date, suggesting that radial patterning that occurs during

embryonic development may influence the post-embryonic pattern generated by the meristematic initials (Scheres et al., 1995, Development 121:53-62).

Characterization of the mutant cell layer in *shr* indicated that two endodermal-specific markers were absent (Benfey et al., 1993, Development 119:57-70). This provided evidence that the wild-type *SHR* gene may be involved in the specification of endodermis identity.

2.3. GEOTROPISM

In plants, the capacity for gravitropism has been correlated with the presence of amyloplast sedimentation. See, *e.g.*, Volkmann and Sievers, 1979, Encyclopedia Plant Physiol., N.S. vol 7, pp. 573-600; Sack, 1991, Intern. Rev. Cytol. 127:193-252; Björkmann, 1992, Adv. Space Res. 12:195-201; Poff et al., in The Physiology of Tropisms, Meyerowitz & Somerville (eds); Cold Spring Harbor Laboratory Press, Plainview, NY (1994) pp. 639-664; Barlow, 1995, Plant Cell Environ. 18:951-962. Amyloplast sedimentation only occurs in cells in specific locations at distinct developmental stages. That is, when and where sedimentation occurs is precisely regulated (Sack, 1991, Intern. Rev. Cytol. 127:193-252). In roots, amyloplast sedimentation only occurs in the central (columella) cells of the rootcap; as these cells mature into peripheral cap cells, the amyloplasts no longer sediment (Sack & Kiss, 1989, Amer. J. Bot. 76:454-464; Sievers & Braun, in The Root Cap: Structure and Function, Wassail et al. (eds.), New York: M. Dekker (1996) pp. 31-49). In stems of many plants, including Arabidopsis, amyloplast sedimentation occurs in the starch sheath (endodermis) especially in elongating regions of the stem (von Guttenberg, Die Physiologischen Scheiden, Handbuch der Pflanzenanatomie; K. Linsbauer (ed.), Berlin: Gebruder Borntraeger, vol. 5 (1943) p. 217; Sack, 1987, Can. J. Bot. 65:1514-1519; Sack, 1991, Intern. Rev. Cytol. 127:193-252; Caspar & Pickard, 1989, Planta 177:185-197; Volkmann et al., 1993,J. Pl. Physiol. 142:710-6).

Gravitropic mutants have been studied for evidence that proves the role of amyloplast sedimentation in gravity sensing. However, many gravitropic mutations affect downstream events such as auxin sensitivity or metabolism (Masson, 1995, BioEssays 17:119-127). Other mutations seem to affect gene products that process information from gravity sensing. For example, the lazy mutants of higher plants and comparable mutants in mosses can clearly sense and respond to gravity, but the mutations reverse the normal polarity of the gravitropic response (Gaiser & Lomax, 1993, Plant Physiol. 102:339-344; Jenkins et al., 1986, Plant Cell Environ 9:637-644). Other mutations appear to affect gravitropism of specific organs. For example, *sgr* mutants have defective shoot gravitropism (Fukaki et al., 1996, Plant Physiol. 110:933-943; Fukaki et al., 1996, Plant Physiol. 110:945-955; Fukaki et al., 1996, Plant Res. 109:129-137).

Citation or identification of any reference herein shall not be construed as an admission that such reference is available as prior art to the present invention.

3. SUMMARY OF THE INVENTION

The structure and function of a regulatory gene, *SCARECROW* (*SCR*), is described. The *SCR* gene is expressed specifically in root progenitor tissues of embryos, and in certain tissues of roots and stems. *SCR* expression controls cell division of certain cell types in roots, and affects the organization of root and stem. The present invention relates to the *SCARECROW* (*SCR*) gene (which encompasses the Arabidopsis *SCR* gene and its orthologs and paralogs), *SCR*-like genes, *SCR* gene products, (including, but not limited to, transcriptional products such as mRNAs, antisense and ribozyme molecules, and translational products such as the SCR protein, polypeptides, peptides and fusion proteins related thereto), antibodies to *SCR* gene products, *SCR* regulatory regions and the use of the foregoing to improve agronomically valuable plants.

The invention is based, in part, on the discovery, identification and cloning of the gene responsible for the *scarecrow* phenotype. In contrast to the prevailing view that the *SCR* gene was likely to be involved in the specification of endodermis, the inventors have determined that the mutant cell layer in roots of *scr* mutants has differentiated characteristics of both cortex and endodermis. This is consistent with a role for *SCR* in the regulation of asymmetric cell division rather than in specification of the identity of either cortex or endodermis. The inventors have determined also that SCR expression affects the gravitropism of plant aerial structures such as the stem.

One aspect of the invention relates to the heterologous expression of *SCR* genes and related nucleotide sequences, and specifically the Arabidopsis *SCR* and maize *ZCARECROW (ZCR)* genes, in stably transformed higher plant species. Modulation of *SCR* and *ZCR* expression levels can be used to advantageously modify root and aerial structures of transgenic plants and enhance the agronomic properties of such plants.

Another aspect of the invention relates to the use of promoters of *SCR* genes, and specifically the use of the Arabidopsis *SCR* and maize *ZCR* promoters to control the expression of protein and RNA products in plants. Plant *SCR* promoters have a variety of uses, including, but not limited to, expressing heterologous genes in the embryo, root, root nodule and stem of transformed plants.

The invention is illustrated by working examples, described *infra,* which demonstrate the isolation of the Arabidopsis *SCR* gene using insertion mutagenesis. More specifically, T-DNA tagging of genomic and cDNA clones of the Arabidopsis *SCR* gene are described. Other working examples include the isolation of *SCR* sequences from plant genomes using PCR amplification in combination with screening of genomic libraries, and heterologous gene expression in transgenic plants using *SCR* promoter expression constructs. Additional working examples describe the cloning and

isolation of maize *ZCR* genes using probes derived from the Arabidopsis *SCR* gene on a maize genomic library. Still other working examples describe the characterization of the maize *ZCR* expression pattern in primary and embryonic roots, and during regeneration of the root tip following excision of the QC.

Structural analysis of the deduced amino acid sequence of Arabidopsis SCR protein indicates that *SCR* encodes a transcription factor. Northern analysis, *in situ* hybridization analysis and enhancer trap analysis show highly localized expression of Arabidopsis *SCR* and maize *ZCR* in embryos and roots. Genetic analysis shows *SCR* expression also affects gravitropism of aerial structures (*e.g.*, stems and shoots). This indicates that *SCR* is also expressed in those structures.

Computer analysis of the deduced amino acid sequence of Arabidopsis SCR protein with those of Expressed Sequence Tag (EST) sequences and genomic sequences in GenBank reveals the existence of at least eighteen *SCR* genes in Arabidopsis, one *SCR* gene in maize, four *SCR* genes in rice, and one *SCR* gene in Brassica. A further aspect of the invention relates to the use of such EST sequences to obtain larger and/or complete clones of the corresponding *SCR* gene.

The various embodiments of the claimed invention presented herein are by way of illustration only and are in no manner intended to limit the scope of the invention.

3.1. DEFINITIONS

As used herein, the terms listed below will have the meanings indicated.

35S = cauliflower mosaic virus promoter for the 35S transcript

cDNA = complementary DNA

cis-regulatory element = A promoter sequence 5' upstream of the TATA box that confers specific regulatory response to a promoter containing such an element. A

promoter may contain one or more cis-regulatory elements, each responsible for a particular regulatory response

coding sequence = sequence that encodes a complete or partial gene product (*e.g.*, a complete protein or a fragment thereof)

DNA = deoxyribonucleic acid

EST = expressed sequence tag

functional portion = a functional portion of a promoter is any portion of a promoter that is capable of causing transcription of a linked gene sequence, *e.g.*, a truncated promoter

gene fusion = a gene construct comprising a promoter operably linked to a heterologous gene, wherein said promoter controls the transcription of the heterologous gene

gene product = the RNA or protein encoded by a gene sequence

gene sequence = sequence that encodes a complete gene product (*e.g.*, a complete protein)

GUS = 1,3-β-Glucuronidase

gDNA = genomic DNA

heterologous gene = In the context of gene constructs, a heterologous gene means that the gene is linked to a promoter that said gene is not naturally linked to. The heterologous gene may or may not be from the organism contributing said promoter. The heterologous gene may encode messenger RNA (mRNA), antisense RNA or ribozymes

homologous promoter = a native promoter of a gene that selectively hybridizes to the sequence of a *SCR* gene described herein

mRNA = messenger RNA

operably

linked	=	A linkage between a promoter and gene sequence such that the transcription of said gene sequence is controlled by said promoter
ortholog	=	related gene in a different plant (*e.g.*, maize *ZCARECROW* gene is an ortholog of the Arabidopsis *SCR* gene)
paralog	=	related gene in the same plant (*e.g.*, Arabidopsis *SCLa1* is a paralog of Arabidopsis *SCR* gene)
RNA	=	ribonucleic acid
RNase	=	ribonuclease
SCR (italic)	=	*SCARECROW* gene or portion thereof, encompasses *SCR* and *ZCR* genes and their orthologs and paralogs
SCR	=	SCARECROW protein
scr (lower case)	=	*scarecrow* mutant (*e.g.*, *scr1*)
SCL	=	*SCARECROW*-like gene
ZCR	=	maize *ZCARECROW* gene, an ortholog of, for example, the Arabidopsis *SCR* gene

SCR protein means a protein containing sequences or a domain substantially similar to one or more motifs (*i.e.*, Motifs I-VI), preferably MOTIF III (VHIID), of the Arabidopsis SCR protein as shown in FIGS. 13A-F and FIGS. 15A-S. SCR proteins include SCR ortholog and paralog proteins having the structure and activities described herein.

SCR polypeptides and peptides include deleted or truncated forms of the SCR protein, and fragments corresponding to the SCR motifs described herein.

SCR fusion proteins encompass proteins in which the SCR protein or an SCR polypeptide or peptide is fused to a heterologous protein, polypeptide or peptide.

SCR gene, nucleotides or coding sequences mean nucleotides, *e.g.*, gDNA or cDNA encoding SCR protein, SCR polypeptides, peptides or SCR fusion proteins.

SCR gene products include transcriptional products such as mRNAs, antisense and ribozyme molecules, as well as translational products of the *SCR* nucleotides described herein, including, but not limited to, the SCR protein, polypeptides, peptides and/or SCR fusion proteins.

SCR promoter means the regulatory region native to the *SCR* gene in a variety of species, which promotes the organ and tissue specific pattern of *SCR* expression described herein.

4. BRIEF DESCRIPTION OF THE FIGURES

FIGS. 1A-B. Schematic of Arabidopsis root anatomy. FIG. 1A. Transverse section showing the four tissues, epidermis, cortex, endodermis and pericycle that surround the vascular tissue. In the longitudinal section, the epidermal/lateral root cap initials and the cortex/endodermal initials are shown at the base of their respective cell files. FIG. 1B. Schematic of division pattern of the cortex/endodermal initial. The initial expands then divides anticlinally to reproduce itself and a daughter cell. The daughter then divides periclinally to produce the progenitors of the endodermis and cortex cell lineages. Abbreviations: C, cortex; Da, daughter cell; E, endodermis; In, initial.

FIGS. 2A-F. Phenotype of *scr* mutant plants. FIG. 2A. Shown left to right are 12-day *scr-2*, *scr-1* and wild-type seedlings grown vertically on nutrient agar medium. FIG. 2B. 21-day *scr-2* mutant plants in soil. FIG. 2C. Transverse section through primary root of 7-day *scr-2*. FIG. 2D. Transverse section through primary root of 7-day wild-type (WT). FIG. 2E. Transverse section through lateral root of 12-day *scr-1* mutant seedling. FIG. 2F. Transverse section through root regenerated from *scr-1* callus. Bar, 50 μm. Abbreviations: C, cortex; En, endodermis; Ep, epidermis; M, mutant cell layer; P, pericycle; V, vascular tissue.

FIGS. 3A-F. Characterization of the cellular identity of the mutant cell layer. FIG. 3A. Endodermis-specific Casparian band staining of transverse sections

through the primary root of 7-day *scr-1* mutant. (Note: the histochemical stain also reveals xylem cells in the vascular cylinder.) FIG. 3B. Casparian band staining of transverse sections through the primary root of 7-day wild-type (WT). FIG. 3C. Immunostaining with the endodermis (and a subset of vascular tissue) specific JIM13 monoclonal antibodies on transverse root sections of *scr-2* mutant. FIG. 3D. Immunostaining with JIM13 monoclonal antibodies on transverse root sections of WT. FIG. 3E. Immunostaining with the JIM7 monoclonal antibody that stains all cell walls on transverse root sections of *scr-2* mutant. FIG. 3F. Immunostaining with JIM7 monoclonal antibodies on transverse root sections of WT. Bar, 25 μm. Abbreviations are same as those for description of FIGS. 2A-2F and: Ca, casparian strip.

FIGS. 4A-F. Immunostaining. FIG. 4A. Immunostaining with the cortex (and epidermis) specific CCRC-M2 monoclonal antibodies on transverse root sections of *scr-1* mutant. FIG. 4B. Immunostaining with CCRC-M2 antibodies on transverse root sections of *scr-2* mutant. FIG. 3C. Immunostaining with CCRC-M2 antibodies on transverse root sections of wild-type (WT). FIG. 4D. Immunostaining with the CCRC-M1 monoclonal antibodies (specific to a cell wall epitope found on all cells) on transverse root sections of *scr-1*. FIG. 4E. Immunostaining with CCRC-M1 antibodies on transverse root sections of *scr-2*. FIG. 4F. Immunostaining with CCRC-M1 antibodies on transverse root sections of WT. Bar, 30 μm. Abbreviations are same as those for description of FIGS. 2A-2F.

FIG. 5A-E. Structure of the Arabidopsis *SCARECROW* gene. FIG. 5A. Nucleic acid sequence and deduced amino acid sequence of the Arabidopsis *SCR* genomic region (SEQ ID NO:1) and (SEQ ID NO:2), respectively. Regulatory sequences including: (i) TATA box, (ii) ATG start codon, and (iii) potential polyadenylation sequence are underlined. Within the deduced amino acid sequence, homopolymeric repeats are underlined. FIG. 5B. Schematic diagram of genomic clone indicating possible functional motifs, T-DNA insertion sites

and subclones used as probes. Abbreviations: Q,S,P,T, region with homopolymeric repeats of these amino acids; b, region with similarity to the basic region of bZIP factors; I and II, regions with leucine heptad repeats; E, acidic region.

FIG. 5C. Comparison of the charged region found in Arabidopsis SCR protein with that found in bZIP transcription factors, SCR bZIP-like domain (SEQ ID NO:3), GCN4 (SEQ ID NO:4), TGA1 (SEQ ID NO:5), C-Fos (SEQ ID NO:6), c-JUN (SEQ ID NO:7), CREB (SEQ ID NO:8), Opaque-2 (SEQ ID NO:9), OBF2 (SEQ ID NO:10), RAF-1 (SEQ ID NO:11). FIG. 5D. Translations of EST clones encoding putative peptide having similarities to the VHIID domain region of Arabidopsis SCR protein (SEQ ID NO:12), F13896 (SEQ ID NO:13), Z37192 (SEQ ID NO:14), and Z25645 (SEQ ID NO:15) are from Arabidopsis, T18310 (SEQ ID NO:17) is from maize and D41474 (SEQ ID NO:16) is from rice. FIG. 5E. The deduced amino acid sequence of the Arabidopsis *SCARECROW* gene (SEQ ID NO:2).

FIGS. 6A-B. Expression of the Arabidopsis *SCARECROW* gene. FIG. 6A. Northern blot of total RNA from wild-type siliques (Si), roots (R), leaves (L) and whole seedlings (Sd) hybridized with Arabidopsis *SCR* probe a and with a probe from the Arabidopsis glutamine dehydrogenase (*GDH*) gene (Melo-Oliveira et al., 1996, Proc. Natl. Acad. Sci. USA 93:4718-4723) as a control for RNA integrity. (*GDH* expression is lower in siliques than in vegetative tissues.) The 1.6 kb band corresponds to the *GDH* gene and the approximately 2.5 kb band corresponds to *SCR*. Ribosomal RNA is shown as a loading control. FIG. 6B. Northern blot of Arabidopsis wild-type, *scr-1* and *scr-2* total RNA, probed with Arabidopsis *SCR* probe "a" corresponding to a cDNA sequence shown in FIG. 5B, and with the *GDH* probe. In *scr-2* mutant additional bands of 4.1 kb and 5.0 kb were detected.

FIGS. 7A-G. *In situ* hybridization and enhancer trap analyses of Arabidopsis *SCR* expression. FIG. 7A. *SCR* RNA expression detected by *in situ* hybridization of *SCR* antisense probe to a longitudinal section through the root meristem. FIG. 7B. *In situ* hybridization of *SCR* antisense

probe to a transverse section in the meristematic region. FIG. 7C. *In situ* hybridization of *SCR* antisense probe to late torpedo stage embryo. FIG. 7D. Negative control *in situ* hybridization using a *SCR* sense probe to a longitudinal section through the root meristem. FIG. 7E. GUS expression in a whole mount in the enhancer trap line, ET199 in primary root tip. FIG. 7F. GUS expression in the ET199 line in transverse root section in the meristematic region. FIG. 7G. GUS expression in ET199 detected in a section through the root meristem. GUS expression is observed in the cortex/endodermal initial, and in the first cell in the endodermal cell lineage but not in the first cell of the cortex lineage. Expression in two endodermal layers is observed higher up in the root because the section was not median at that point. Bar, 50 μm. Abbreviations are same as those in the description of FIGS. 2A-2F.

FIG. 8. Partial nucleotide sequence (SEQ ID NO:18) and deduced amino acid sequence (SEQ ID NO:19) of the Arabidopsis *SCLa4* gene.

FIG. 9. Partial nucleotide sequence (SEQ ID NO:20) and deduced amino acid sequence (SEQ ID NO:21) of the Arabidopsis *SCLa3* gene.

FIG. 10. Partial nucleotide sequence (SEQ ID NO:22) of the Arabidopsis *SCLa1* gene.

FIG. 11A. Nucleotide sequence (SEQ ID NO:24) and deduced amino acid sequence (SEQ ID NO:25) of the maize Zm-Scl1 fragment.

FIG. 11B. Partial nucleotide sequence (SEQ ID NO:25) and deduced amino acid sequence (SEQ ID NO:26) of the maize *SCLm1* gene (Zm-Scl2).

FIG. 12A-B. Nucleotide sequence of rice *SCLo3* EST clone. FIG. 12A. Sequence of 5' end of EST clone (SEQ ID NO:28). FIG. 12B. Sequence of 3' end of EST clone (SEQ ID NO:29).

FIGS. 13A-F. Comparison of the amino acid sequence of members of the *SCARECROW* family of genes. Conserved Motifs I through VI are indicated by dashed line above the

aligned sequences. Consensus sequences are shown in bold. See Table 1 for the identity and sequence identifier number of each of the sequences shown in this Figure.

FIG. 14. Restriction map of the approximately 8.8 kb Eco RI insert DNA of lambda clone, t643, containing the Arabidopsis *SCR* gene. The locations of the approximately 5.6 kb HindIII-SacI fragment subcloned in plasmid LIG 1-3/SAC+MoB_2 1SAC, and the *SCR* coding region are indicated below the restriction map. The location of the translational initiation site of the *SCR* gene is at the Nco I site at the left end of the indicated coding region. The *SCR* coding sequence begins at the translation initiation site and extends approximately 1955 nucleotides to its right. *E. coli* DH5α containing plasmid pLIG1-3/SAC+MoB_2 1SAC, has the ATCC accession number 98031.

FIGS. 15A-S. Comparison of the partial and complete amino acid sequences of several plant members of the *SCARECROW* family of genes. The amino acid sequences are aligned in a manner that maximizes amino acid sequence similarity and identity among *SCR* family members. Each sequence shown is continuous except where noted otherwise; the dots are inserted between two sequence segments in order to align homologous segments. "X" in the middle of a sequence indicates ambiguity in the corresponding nucleotide sequence and, possible termination of the ORF at the "X" residue site. "X" at the end of a sequence indicates termination of the ORF at the "X" residue site. The numbering of the amino acid residues is shown at the bottom of each figure and is based on the Arabidopsis SCR amino acid sequence. Conserved Motifs I through VI are indicated by the various dashed lines above the figures. The new and old names of the family members are shown in FIG. 15A. The sequences of *SCR*, Tf1 and Tf4 are of the complete SCR protein. See Table 1 for the identity and the sequence identifier number of each sequence shown in these figures.

FIGS. 16A-M. The partial nucleotide sequences of several plant members of the *SCARECROW* family of genes. "N"

indicates an unknown base. See Table 1 for the identity and the sequence identifier number of each sequence shown in these figures.

FIG. 17A. The partial nucleotide sequence (SEQ ID NO:66) of the maize *ZCR* gene.

FIG. 17B. The partial amino acid sequence (SEQ ID NO:67) of the maize *ZCR* gene. The underlined sequence shares approximately 80% sequence identity with a corresponding sequence of Arabidopsis SCR protein.

FIG. 18. Comparison of the partial amino acid sequences of several *SCR* ortholog sequences amplified from the genomes of carrot, soybean and spruce. The *SCLd1* and *SCLp1* sequences each were obtained by PCR amplification using a combination of 1F and 1R primers. The *SCLg1* sequence was obtained by PCR amplification using a combination of 1F and WP primers. *See*, for example, Section 5.1.1., *infra*. The amino acid sequences are aligned in a manner that maximizes amino acid sequence identity and similarity amongst these sequences. Each sequence shown is continuous except where noted otherwise; the dashes are inserted between two sequence segments in order to allow alignment of homologous segments. "x" in the middle of a sequence indicates ambiguity in the corresponding nucleotide sequence and, possible termination of the ORF or existence of an intron at the "x" residue site. See Table 1 for the identity and the sequence identifier number of each sequence shown in this figure.

FIG. 19. Comparison of promoter activities in transgenic lines and roots. **Panel a.** A stably transformed line containing four copies of the B2 subdomain of the 35S promoter of CaMV upstream of GUS (Benfey et al., 1990). GUS is expressed in the root tip. **Panel b.** Roots emerging from callus transformed with four copies of the B2 subdomain of the 35S promoter fused to GUS. GUS expression can be seen in the emerging root tips (arrows). **Panel c.** Higher magnification of a root emerging from the callus in panel b. GUS is clearly restricted to the root tip. The morphology of roots regenerated from calli often appears abnormal. **Panel**

d. A transgenic plant regenerated from the calli and roots shown in panel b. GUS expression in this plants appears to be similar to that of the original line shown in panel a. **Panel e.** ET199, a stably transformed line that contains an enhancer trapping construct with a minimal promoter fused to the GUS coding region inserted 1 kb upstream from the *SCR* coding region. GUS expression is primarily in the endodermal layer of the root. **Panel f.** Roots emerging from calli transformed with the *SCR* promoter::GUS construct. Expression of the GUS gene appears to be limited to an internal layer (arrows). **Panel g.** *SCR* promoter::GUS transformed root in liquid culture. Roots shown in panel f were excised and transferred to liquid cultures. GUS expression is primarily found in the endodermal layer as in ET199. The expression of GUS in the quiescent center, as seen here, is also sometimes observed in ET199. Bar, 50μm.

FIG. 20. Analysis of *SCR* promoter activity in the *scr* mutant background. **Panel a.** Roots emerging from *scr* calli transformed with the *SCR* promoter::GUS construct. Roots regenerated from *scr* calli are very short. GUS expression appears to be limited to an internal layer of the root (arrows). **Panel b.** Root regenerated from transformed *scr* calli and transferred to liquid culture. The *scr* phenotype, a single layer between the epidermis and pericycle, is easily seen. GUS expression is limited to this mutant layer. E, Epidermis. M, Mutant Layer. P, Pericycle. Bar, 50μm.

FIG. 21. Molecular Complementation of the *scr* mutant. **Panels a, c and e.** *scr* transformed with the *SCR* promoter::GUS construct. **Panels b, d and f.** *scr* transformed with the *SCR* promoter::*SCR* coding region construct. **Panels a and b.** Roots emerging from *scr* calli. Arrows point to several very short roots among many fine root hairs in the *scr* calli transformed with the *SCR* promoter::GUS construct. In contrast, roots from *scr* calli transformed with the *SCR* promoter::*SCR* coding region construct appeared to be wild-type in length, suggesting molecular complementation by

the transgene. **Panels c and d.** Transgenic roots in liquid culture. The *scr* roots transformed with the *SCR* promoter::GUS construct appeared short, while those transformed with the *SCR* promoter::*SCR* coding region construct appeared of wild-type length. **Panels e and f.** Transverse sections through roots emerging from calli. Whereas there is only a single cell layer between the epidermis and stele in the *SCR* promoter::GUS transformed root, the radial organization of the root transformed with the *SCR* promoter::*SCR* coding region appeared identical to wild-type, with both cortex and endodermal layers. E, epidermis. M, mutant layer. C, cortex. En, Endodermis. P, Pericycle. Bar, 50μm

FIGS. 22A-F. Expression of *ZCR* in maize root tips. FIG. 22A. Expression of *ZCR* is in the endodermal layer and extends down through the region of the quiescent center. FIGS. 22B-C. Higher magnification showing expression in a single cell layer through the quiescent center. FIG. 22D. Expression of *ZCR* in the maize embryonic root. FIG. 22E. Higher magnification showing expression in the embryonic root. FIG. 22F. Expression of *ZCR* in the maize lateral root.

FIGS. 23 A-B. Root apical meristems of maize and Arabidopsis. Both show a type of a closed meristem in which all files of cells converge onto a pole at the root apex, making the boundary between the root proper and the root cap discrete. FIG. 23A. A schematic representation of the monocotyledonous closed-type of root apical meristem of maize. FIG. 23B. A schematic representation of the dicotyledonous closed-type of root apical meristem of Arabidopsis.

FIGS. 24A-G. Embryo development in Maize. FIG. 24A. Three-celled embryo establishing the initial asymmetry and showing the first division of a terminal cell. FIGS. 24B-C. Embryos showing embryo proper and suspensor. FIGS. 24D-E. Embryos showing radial asymmetry and the initial development of shoot and root apical meristems.

FIGS. 24F-G. Embryos showing the elaborate organization of shoot and root apical meristems.

FIG. 25. Maize Scarecrow gene. The nucleotide and deduced amino acid sequnce of the maize scarecrow gene (*ZCR*) is shown. The amino acid numbers are shown on the right, while the nucleotides are numbered on the left.

FIG. 26. Amino acid sequence alignment of maize *ZCR* and Arabidopsis *SCR*. Identical residues are marked by asterisks. In addition, three copies of an LXXLL motif are underlined.

FIGS. 27A-G. Maize Scarecrow gene expression during regeneration of the root apex following excision of the QC. FIGS. 27A-B. Immediately after removal of the root cap and excision of the QC, no significant alteration in the expression pattern was observed. FIGS. 27C-D. Maize expression pattern 24 hours following excision of the QC. These figures show isolated expression of the gene between cell files. FIG. 27E. Expression 48 hours following excision of the QC. This figure shows that the root tip has regained much of its normal shape, although the cell files have not organized into the converging files seen in normal roots. FIG. 27F. Expression 72 hours following excision of the QC. At this stage, the expression pattern resembles that found in the unexcised root. FIG. 27G. Expression 96 and 100 hours following excission of the QC. At this stage, the expression pattern is similar to that in the primary root.

FIGS. 28A-AH. The partial nucleotide and amino acid sequences of Arabidopsis EST's that encode members of the *SCARECROW*-like (*SCL*) gene family. "N" indicates an unknown base. See Table 2 for the identity and the sequence identifier number of each sequence shown in these figures.

FIG. 29. Alignment of the Arabidopsis GRAS gene products. The highly conserved region of the GRAS products can be divided into five recognizable motifs, indicated in the figure. *See also*, for example, Section 5.1.5., *infra*. The absolutely conserved residues within the VHIID and SAW motifs are highlighted in bold, as are the hydrophobic

residues of the leucine heptads, the P-F-Y-R-E residues of the PFYRE motif, and the two short sequences that define the end of the VHIID motif and the beginning of the PFYRE motif. The @ symbol in the alignment indicates the location of an apparent insertion in the *SCL3* gene. The deduced amino acid sequence of the insertion is shown at the bottom of the figure.

FIG. 30. RNA Gel Blot. mRNA from siliques (Si) and 14 day old shoots (Sh) and roots (R) was isolated and analyzed by RNA gel blot hybridization with specific antisense digoxygenin-labeled probes. The *SCL*s analyzed are all expressed within the roots, and many of them are expressed in all of the organs tested. As the amount of mRNA loaded on the gels and the exposure times for all of these blots varied, direct comparisons of the levels of expression are not possible. Detection of *SCL1*, however, required significantly shorter exposures than the others, and *SCL6*, *SCL7* and *SCL9* required significantly longer exposures and more mRNA. A representative ethidium bromide-stained RNA gel is shown below as a loading control.

FIG. 31. *In situ* Hybridizations with *SCR* and *SCL3*. Transverse sections (a, b, and d) and a longitudinal section (c) of 7 day old roots were hybridized with either an antisense *SCR* riboprobe (a), an antisense *SCL3* riboprobe (b and c) or a sense *SCL3* riboprobe (d). Strong signal is observed in the endodermis with the antisense *SCR* probe and the antisense *SCL3* probe, but not with the sense *SCL3* probe. Scale bars in (a) and (c) are both 25 mm. The magnification is the same in panels (a), (b), and (d).

FIG. 32. RNA Blot Analysis. An RNA blot analysis in which either total RNA or poly-A selected RNA from roots (R) and shoots (S) were probed with the full-length *ZCR* cDNA. The hybridizing band is approximately 2.6 kilobases.

FIG. 33. CBPBTT44 Partial cDNA and Amino Acid Sequence. The partial nucleotide and amino acid sequence of CBPBTT44, a closely related gene to the maize *ZCR* gene.

FIG. 34. Alignment of the Arabidopsis *SCR*, the maize *ZCR* and the CBPBTT44 amino acid sequence. As shown in bold, all three genes contain the leucine heptad repeats. The alignment further shows that all three genes share a high degree of homology.

FIG. 35. Southern Blot Analysis. A Southern of maize genomic DNA probed with (left) the maize *ZCR* cDNA, wherein the "H" lane represents DNA digested with HindIII and the "RV" lane represents DNA digested with EcoRV restriction enzymes; (right) gene-specific probes (A) maize *ZCR* cDNA for comparison; (B) maize *ZCR* gene-specific probe and (C) CBPBTT44 gene-specific probe. The results demonstrate that CBPBTT44 is the source of the other hybridizing bands picked up by the maize *ZCR* cDNA.

5. DETAILED DESCRIPTION OF THE INVENTION

The invention relates to the *SCARECROW* (*SCR*) gene; *SCARECROW*-like (*SCL*) genes, *SCR* gene products, including, but not limited to, transcriptional products such as mRNAs, antisense and ribozyme molecules, translational products such as the SCR protein, polypeptides, peptides and fusion proteins related thereto; antibodies to *SCR* gene products; *SCR* regulatory regions; and the use of the foregoing to improve agronomically valuable plants.

In summary, the data described herein show the identification of *SCR*, a gene involved in the regulation of a specific asymmetric division, in controlling gravitropic response in aerial structures, and in controlling pattern formation in roots. Sequence analysis shows that the SCR protein has many hallmarks of transcription factors. *In situ* and marker line expression studies show that *SCR* is expressed in the cortex/endodermal initial of roots before asymmetric division occurs, and in the quiescent center of regenerating roots. Together, these findings indicate that the *SCR* gene regulates key events that establish the asymmetric division that generates separate cortex and endodermal cell lineages,

and that affect tissue organization of roots. The establishment of these lineages is not required for cell differentiation to occur, because in the absence of division, the resulting cell acquires mature characteristics of both cortex and endodermal cells. However, it is possible that *SCR* functions to establish the polarity of the initial before cell division, or that it is involved in generating an external polarity that has an effect on asymmetric cell division.

Genetic analysis indicates that *SCR* expression affects gravitropism of plant stems, hypocotyls and shoots. This indicates that *SCR* is expressed also in these aerial structures of plants.

The *SCR* genes and promoters of the present invention have a number of important agricultural uses. The *SCR* promoters of the invention may be used in expression constructs to express desired heterologous gene products in the embryo, root, root nodule, and starch sheath layer in the stem of transgenic plants transformed with such constructs. For example, *SCR* promoters may be used to express disease resistance genes such as lysozymes, cecropins, maganins or thionins for anti-bacterial protection, or the pathogenesis-related (PR) proteins such as glucanases and chitinases for anti-fungal protection. *SCR* promoters also may be used to express a variety of pest resistance genes in the aforementioned plant structures and tissues. Examples of useful gene products for controlling nematodes or insects include *Bacillus thuringiensis* endotoxins, protease inhibitors, collagenases, chitinase, glucanases, lectins and glycosidases.

Gene constructs that express or ectopically express *SCR*, and the *SCR*-suppression constructs of the invention, may be used to alter the root and/or stem structure, and the gravitropism of aerial structures of transgenic plants. Since *SCR* regulates root cell divisions, overexpression of *SCR* can be used to increase division of certain cells in roots and thereby form thicker and stronger roots. Thicker

and stronger roots are beneficial in preventing plant lodging. Conversely, suppression of *SCR* expression can be used to decrease cell division in roots and thereby form thinner roots. Thinner roots are more efficient in uptake of soil nutrients. Since *SCR* affects gravitropism of aerial structures, overexpression of *SCR* may be used to develop "straighter" transgenic plants that are less susceptible to lodging.

Further, the *SCR* gene sequence may be used as a molecular marker for a quantitative trait, *e.g.*, a root or gravitropism trait, in molecular breeding of crop plants.

For purposes of clarity and not by way of limitation, the invention is described in the subsections below in terms of (a) *SCR* genes and nucleotides; (b) *SCR* gene products; (c) antibodies to *SCR* gene products; (d) *SCR* promoters and promoter elements; (e) transgenic plants which ectopically express *SCR*; (f) transgenic plants in which endogenous *SCR* expression is suppressed; and (g) transgenic plants in which expression of a transgene of interest is controlled by the *SCR* promoter.

5.1. *SCR* GENES

The *SCARECROW* genes and nucleotide sequences of the invention include: (a) a gene listed below in Tables 1 or 2 (hereinafter, a gene comprising any one of the nucleotide sequences shown in FIG. 5A, FIG. 8, FIG. 9, FIG. 10, FIGS. 11A-B, FIGS. 12A-B, FIGS. 16A-M, FIG. 17A, FIG. 25 or FIGS. 28A-AH, or a segment of such nucleotide sequences), or as contained in the clones described herein and deposited with the ATCC (see Section 13, infra); (b) a nucleotide sequence that encodes a protein comprising any one of the amino acid sequences shown in FIG. 5A, FIG. 5D, FIG. 5E, FIG. 8, FIG. 9, FIGS. 11A-B, FIGS. 13A-F, FIGS. 15A-S, FIG. 17B, FIG. 18 or FIG. 25, or a segment of such amino acid sequences, or that is encoded by any one of the genes and/or nucleotide sequences listed by their sequence identifier numbers in Tables 1 or 2, or any segment of such genes and/or nucleotide

sequences, or contained in any one of the clones described herein and deposited with the ATCC (see Section 13, infra); (c) any gene comprising a nucleotide sequence that hybridizes to the complement of any one of the genes and/or nucleotide sequences listed by their sequence identifier numbers in Tables 1 or 2, or any segment of such genes and/or nucleotide sequences, or as contained in any one of the clones described herein and deposited with the ATCC, under highly stringent conditions, *e.g.*, hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1xSSC/0.1% SDS at 68°C (Ausubel F.M. et al., eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York, at p. 2.10.3) and that encodes a gene product functionally equivalent to *SCR* gene product encoded completely or partly by any one of the genes and/or sequences listed in Tables 1 or 2 or any segment of such genes and nucleotide sequences, or as contained in any one of the clones deposited with the ATCC; (d) any gene comprising a nucleotide sequence that hybridizes to the complement of any one of the sequences listed by their sequence identifier numbers in Tables 1 or 2, or any segment of such nucleotide sequences, or as contained in any one of the clones described herein and deposited with the ATCC, under less stringent conditions, such as moderately stringent conditions, *e.g.*, washing in 0.2xSSC/0.1% SDS at 42°C (Ausubel et al., 1989, *supra*), and which encodes a functionally equivalent *SCR* gene product; (e) any gene comprising a nucleotide sequence that hybridizes to the complement of any one of the sequences listed by their sequence identifier numbers in Tables 1 or 2 or any segment of such nucleotide sequences, or as contained in any one of the clones described herein and deposited with the ATCC, under the following low stringency conditions: pre-hybridization in hybridization solution (HS) containing 43% formamide, 5xSSC, 1% SDS, 10% dextran sulfate, 0.1% sarkosyl, 2% block (Genius kit, Boehringer-Mannheim), followed by hybridization overnight at 30 to 33°C using as a probe a DNA

molecule of approximately 1.6 kb of SEQ ID NO:1 at a concentration of 20 ng/ml, followed by washing in 2xSSC/0.1% SDS two times for 15 minutes at room temperature and then two times at 50°C, and which encodes a functionally equivalent *SCR* gene product; and/or (f) any gene comprising a nucleotide sequence that encodes a polypeptide or protein containing the consensus sequence for SCR (*i.e.*, MOTIF III or VHIID) shown in FIGS. 13B-D or a segment of such polypeptide or protein. The partial and complete nucleotide and amino acid sequences of SCR genes and encoded proteins and polypeptides included in the invention are listed in Tables 1 or 2 below.

Table 1. *SCR* ORTHOLOGS AND PARALOGS

New Name	Old Name	EST Clone[1]	SEQ ID NOs Nucleotide[3]	Amino Acid
ARABIDOPSIS				
SCLa1	1110	Z25645/33772	22	23
SCLa2	Tf4	Z34599	--	35*
SCLa3	3935	Z37192/1 N96166	20	21
SCLa4	4818	F13896/7	18	19
SCLa5	4871	F13949	45	46
SCLa6	12398	R29793	51	52
SCLa7	3635	T21627 H76979 N96767	55	56
SCLa8	Tf1	T46205 (9468) N96653 (21711)	--	34*
SCLa9	10964	T78186 T44774	47	48
SCLa10	11261	T76483	49	50
SCLa11	18652	N37425	53	54
SCLa12	23196	W43803 W435138 AA042397	57	58
SCLa13	33/08	T46008	--	41
SCR	*Scr*	N.A.[2]	1*	2*
RICE				
SCLo1	713	D15490	--	43
SCLo2	2504	D40482 D40607 D40800 D41389	--	44
SCLo3	3989	D41474	--	36
SCLo4	11846	C20324	--	59
MAIZE				
ZCR	N.A.	N.A.	?	?
SCLm1	18310	T18310	--	37
BRASSICA				

Table 1. (Continued)

New Name	Old Name	EST Clone[1]	SEQ ID NOs Nucleotide[3]	Amino Acid
SCLb1	174	H74669	--	42
CARROT				
SCLd1	N.A.	N.A.	60	61
SOYBEAN				
SCLg1	N.A.	N.A.	62	63
SPRUCE				
SCLp1	N.A.	N.A.	64	65

[1] Each EST clone is identified by its GenBank accession number. Each EST clone corresponds to a deposit of a cDNA sequence that matches a part of the nucleotide sequence of the corresponding *SCR* ortholog or paralog.

[2] N.A. = not applicable.

[3] The partial or complete nucleotide sequence of the *SCR* orthologs and paralogs listed here are shown in FIGS. 5A, 8, 9, 10, 11A-B, 12A-B, 16A-M, 17A and 25.

[+] Contains the complete coding sequence of Arabidopsis *SCR* gene.

[*] Contains the complete amino acid sequence of Arabidopsis *SCLa2*, *SCLa8*, or *SCR* protein.

Designation	Accession Numbers	Accession Number Complete EST Sequence	Map Position
SCL1	Z25645/33772, B10318, B11686	AF0360300	1: m235-g3829 (RI)
GAI	Z34183, Z34599, T22782, Y11337, Y15193, B62171		1: ve006-ve007 (CIC3G6, 4H9, and 11C3)
SCL3	Z37192/Z37191, N96166, B20233, B18969	AF0360301	1: m213 (CIC 1G8, 4H4, 8G4)
SCL4	Z46550, Z38048, Z38085, B22400, B23696 G: AB010700		5 (genomic clone)
SCL5	F13896/F13897, AA395075	AF0360302	1: m213 (RI)
SCL6	F13949 G:AC004708, (WASHU003)	AF0360303	4: mi51 (CIC 2C7, 5B11, 5C11, 10C8) (genomic clone)
SCL7	R29793	AF0360304	3: CDs4, m457 (CIC 8E2, 8E1, 9D1)
SCL8	T21627, H76979, N96767, T43670 AA395639, B77404	AF0360305	5: PAP003 (CIC 11F10)
SCL9	T76186, T44774 G:AC004684, B25776	AF0360306	2: ve018-nga168 (CIC 10F12) (genomic clone)
RGA	T45793, T46205, N96653, Y11336, Y15194		2: ve012 (CIC7C11, 2F4, and 6G2)
SCL11	T76483, AA394557, AA605493	AF0360307	NP
SCL12	F15146		
SCL13 (VHS4)	F15454, N37425, AA720344, R29917 G: Z97343	AF0360308	4: g4539-mi112 (CIC 4D3, 6G4, 2B8, 5E12, 7G8, 12B9) (genomic clone)
SCL14	W43803, W43538, AA042397	AF0360309	NP
SCL15 (VHS5)	N65163 G: Z99708		4 (genomic clone)
SCL16	G: AB007645		5 (genomic clone)
RGL	AJ224957		
SCL18	B10115, B30030, G:AC002328		1: mi209,nga280,nga128 (BAC F20N2) (genomic clone)
SCL19	Z26055, B62171, B62460		
SCR	U62798		3: ve042-ve022 (CIC 11G5, 9D7)

Table 2

Functional equivalents of the *SCR* gene product include any plant gene product that regulates plant embryo or root development, or, preferably, that regulates root cell division or root tissue organization, or affects gravitropism of plant aerial structures (*e.g.*, stems and hypocotyls). Functional equivalents of the *SCR* gene product include naturally occurring *SCR* gene products, and mutant *SCR* gene products, whether naturally occurring or engineered.

The invention also includes nucleic acid molecules, preferably DNA molecules, that hybridize to, and are therefore the complements of the nucleotide sequences (a) through (f), in the first paragraph of this section. Such hybridization conditions may be highly stringent, less highly stringent, or low stringency as described above. In instances wherein the nucleic acid molecules are oligonucleotides ("oligos"), highly stringent conditions may refer, *e.g.*, to washing in 6xSSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligos), 48°C (for 17-base oligos), 55°C (for 20-base oligos), and 60°C (for 23-base oligos). These nucleic acid molecules may act as *SCR* antisense molecules, useful, for example, in *SCR* gene regulation and/or as antisense primers in amplification reactions of *SCR* gene and/or nucleic acid sequences. Further, such sequences may be used as part of ribozyme and/or triple helix sequences, also useful for *SCR* gene regulation. Still further, such molecules may be used as components in probing methods whereby the presence of a *SCARECROW* allele may be detected.

The invention also includes nucleic acid molecules, preferably DNA molecules, which are amplified using the polymerase chain reaction under conditions described in Section 5.1.1., *infra*, and that encode a gene product functionally equivalent to a *SCR* gene product encoded by any one of the genes and sequences listed in Tables 1 or 2 or as contained in any one of the clones described herein and deposited with the ATCC.

The invention also encompasses (a) DNA vectors that contain any of the foregoing gene and/or coding sequences

and/or their complements (*i.e.*, antisense or ribozyme molecules); (b) DNA expression vectors that contain any of the foregoing gene and/or coding sequences operatively associated with a regulatory element that directs the expression of the gene and/or coding sequences; and (c) genetically engineered host cells that contain any of the foregoing gene and/or coding sequences operatively associated with a regulatory element that directs the expression of the gene and/or coding sequences in the host cell. As used herein, regulatory elements include, but are not limited to, inducible and non-inducible promoters, enhancers, operators and other elements known to those skilled in the art that drive and regulate expression.

The invention also encompasses nucleotide sequences that encode mutant *SCR* gene products, peptide fragments of the *SCR* gene product, truncated *SCR* gene products and SCR fusion proteins. These gene products include, but are not limited to, nucleotide sequences encoding mutant *SCR* gene products; polypeptides or peptides corresponding to one or more of the Motifs I-VI as shown in FIGS. 13A-F and FIGS. 15A-S, or the bZIP, VHIID, or leucine heptad domains of the SCR, or portions of these motifs and domains; truncated *SCR* gene products in which one or more of the motifs or domains is deleted, *e.g.*, a truncated, nonfunctional SCR lacking all or a portion of the Motifs I-VI as shown in FIGS. 13A-F and FIGS. 15A-S, or the bZIP, VHIID, or leucine heptad domains of the SCR. Nucleotides encoding fusion proteins may include, but are not limited to, full length SCR, truncated SCR or peptide fragments of SCR fused to an unrelated protein or peptide, such as, for example, an enzyme, fluorescent protein or luminescent protein which can be used as a marker.

In particular, the invention includes, for example, fragments of *SCR* genes encoding one or more of the following

domains as shown in FIG. 5E: amino acids 1-264, 265-283, 287-316, 410-473, 436-473, and 473-653.

In addition to the gene and/or coding sequences described above, homologous *SCR* genes, and other genes related by DNA sequence, may be identified and may be readily isolated, without undue experimentation, by molecular biological techniques well known in the art. More specifically, such homologs include, for example, paralogs (*i.e.*, members of the *SCR* gene family occurring in the same plant) as well as orthologs (*i.e.*, members of the *SCR* gene family which occur in a different plant species) of the Arabidopsis *SCR* gene.

A specific embodiment of a *SCR* gene and coding sequence of the invention is Arabidopsis *SCR* (FIGS. 5A and 5E). Other specific embodiments include the various *SCR* genes and coding sequences listed in Tables 1 or 2, *supra.*

Methods for isolating *SCR* genes and coding sequences are described in detail in Section 5.2, below.

SCR genes share substantial amino acid sequence similarities at the protein level and nucleotide sequence similarities in their encoding genes. The term "substantially similar" or "substantial similarity" when used herein with respect to two amino acid sequences means that the two sequences have at least 75% identical residues, preferably at least 85% identical residues and most preferably at least 95% identical residues. The same term when used herein with respect to two nucleotide sequences means that the two sequences have at least 70% identical residues, preferably at least 85% identical residues and most preferably at least 95% identical residues. Determining whether two sequences are substantially similar may be carried out using any methodologies known to one skilled in the art, preferably using computer assisted analysis. For example, the alignments shown herein were initially accomplished by a BLAST search (NCBI using the BLAST network server). The final alignments of *SCR* family members were done manually.

Moreover, *SCR* genes show highly localized expression in embryos and, particularly, roots. Such expression patterns may be ascertained by Northern hybridizations and *in situ* hybridizations using antisense probes.

5.1.1. ISOLATION OF *SCR* GENES

The following methods can be used to obtain *SCR* and *SCL* genes and coding sequences from a wide variety of plants, including, but not limited to, *Arabidopsis thaliana, Zea mays, Nicotiana tabacum, Daucus carota, Oryza, Glycine max, Lemna gibba* and *Picea abies*.

Nucleotide sequences encoding an *SCR* gene, an *SCL* gene or portions thereof may be obtained by PCR amplification of plant genomic DNA or cDNA. Useful cDNA sources include "free" cDNA preparations (*i.e.*, the products of cDNA synthesis) and cloned cDNA in cDNA libraries. Root cDNA preparations or libraries are particularly preferred.

The amplification may use, as the 5'-primer (*i.e.*, forward primer), a degenerate oligonucleotide that corresponds to a segment of a known SCR amino acid sequence, preferably from the amino-terminal region. The 3'-primer (*i.e.*, reverse primer) may be a degenerate oligonucleotide that corresponds to a distal segment of the same known SCR amino acid sequence (*i.e.*, carboxyl to the sequence that corresponds to the 5'-primer). For example, the amino acid sequence of the Arabidopsis SCR protein (SEQ ID NO:2) may be used to design useful 5' and 3' primers. Preferably, the primers corresponds to segments in the Motif III or VHIID domain of SCR protein (see FIGS. 13B-D and FIGS. 15K-L). The sequence of the optimal degenerate oligonucleotide probe corresponding to a known amino acid sequence may be determined by standard algorithms known in the art. See for example, Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Vol 2 (1989).

Further, for amplification from cDNA sources, the 3'-primer may be an oligonucleotide comprising an 3' oligo(dT) sequence. The amplification also may use as primers nucleotide sequences of *SCR* and *SCL* genes or coding sequences (*e.g.*, any one of the *scr* sequences and EST sequences listed in Table 1 and Table 2).

PCR amplification can be carried out, *e.g.*, by use of a Perkin-Elmer Cetus thermal cycler and Taq polymerase (Gene Amp™). One can choose to synthesize several different degenerate primers for use in the PCR reactions. It also is possible to vary the stringency of hybridization conditions used in priming the PCR reactions, to allow for greater or lesser degrees of nucleotide sequence similarity between the degenerate primers and the corresponding sequences in the cDNA library. One of ordinary skill in the art will know that the appropriate amplification conditions and parameters depend, in part, on the length and base composition of the primers and that such conditions may be determined using standard formulae. Protocols for executing all PCR procedures discussed herein are well known to those skilled in the art, and may be found in references such as Gelfand, 1989, PCR Technology, Principles and Applications for DNA Amplification, H.A. Erlich, ed., Stockton Press, New York; and Current Protocols In Molecular Biology, Vol. 2, Ch. 15, Ausubel et al., eds 1988, New York, Wiley & Sons, Inc.

A PCR amplified sequence may be molecularly cloned and sequenced. The amplified sequence may be utilized as a probe to isolate genomic or cDNA clones of a *SCR* gene, as described below. This, in turn, will permit the determination of a *SCR* gene's complete nucleotide sequence, including its promoter, the analysis of its expression, and the production of its encoded protein, as described *infra*.

In a preferred embodiment, PCR amplification of *SCR* gene and/or coding sequences can be carried out according to the following procedure:

PRIMERS:

Forward:

Name: SCR5AII (23-mer, 2 inosines, 64-mix)
A.A. code: HFTANQAI
DNA Sequence: 5' CAT/C TTT/C ACI GCI AAT/C CAA/G GCN AT 3'

Name: SCR5B (29-mer, 1 inosine, 144-mix)
A.A. code: VHIID(L/F)D
DNA Sequence: 5' ACGTCTCGA GTI CAT/C ATA/C/T ATA/C/T GAT/C TTN GA 3'

Name: 1F
A.A. code: LQCAEAV
DNA Sequence: (T/C)TI CA(A/G) TG(T/C GCI GA(A/G) GCN GT

Reverse:

Name: SCR3AII (23-mer, 2 inosines, 128-mix)
A.A. code: PGGPP(H/N/K)(V/L/F)R'
DNA Sequence: 5' CG/T CCA/C GTG/T TGG IGG ICC NCC NGG 3'

Name: 1R
A.A. code: AFQVFNGI
DNA Sequence: AT ICC (A/G)TT (A/G)AA IAC (C/T)TG (A/G)AA NGC

Name: 4R
A.A. code: QWPGLFHI
DNA Sequence: AT (A/G)TG (A/G)AA IA(A/G) NCC IGG CCA (C/T)TG

I = inosine
N = A/C/G/T

Useful primer combinations include the following: SCR5AII+SCR3AII; SCR5B+SCR3AII; IF+IR; and IF+4R

PCR:

Reaction mixture (volume 50 µl):

-5 µl 10X amplification buffer containing Mg (Boehringer-Mannheim)
-1 µl 10 mM dNTP's
-1 µl forward primer (stock concentration: 80 pmol/µl)
-1 µl reverse primer (80 pmol/µl)
-DNA (100-300 ng).

Begin reaction with "hot start" in which the enzyme is added to the mix only after a brief denaturation at a high temperature (80°C)

Cycles:

94°C 30 sec - brief denaturation (to prevent non-specific priming)
80°C 5 min - apply the enzyme to the tubes (30 tubes/round at maximum)
94°C 5 min - thorough denaturation
2 times: 94°C 1 min
64°C 5 min
72°C 2 min
2 times: 94°C 1 min
62°C 5 min
72°C 2 min
2 times: 94°C 1 min
60°C 5 min
72°C 2 min

(reduce the annealing temperature 2°C in every second round), until 44°C is reached after that:

40 times: 94°C 20 sec
48°C 1 min
72°C 2 min

finally, let cool down to 15°C.

An *SCR* or *SCL* gene coding sequence also may be isolated by screening a plant genomic or cDNA library using an *SCR* or *SCL* nucleotide sequence (*e.g.*, the sequence of any of the *SCR* or *SCL* genes and sequences and EST clone sequences listed in Table 1 and Table 2.) as a hybridization probe. For example, the whole, or a segment, of the Arabidopsis *SCR* nucleotide sequence (FIG. 5A) may be used. Alternatively, a *SCR* or *SCL* gene may be isolated from such libraries using a degenerate oligonucleotide that corresponds to a segment of a SCR amino acid sequence as a probe. For example, a degenerate oligonucleotide probe corresponding to a segment of the Arabidopsis SCR amino acid sequence (FIG. 5E) may be used.

In preparation of cDNA libraries, total RNA is isolated from plant tissues, preferably roots. Poly(A)+ RNA is isolated from the total RNA, and cDNA prepared from the poly(A)+ RNA, all using standard procedures. See, for example, Sambrook et al., Molecular Cloning: A Laboratory Manual, 2d ed., Vol. 2 (1989). The cDNAs may be synthesized with a restriction enzyme site at their 3'-ends by using an

appropriate primer and further have linkers or adaptors attached at their 5'-ends to facilitate the insertion of the cDNAs into suitable cDNA cloning vectors. Alternatively, adaptors or linkers may be attached to the cDNAs after the completion of cDNA synthesis.

In preparation of genomic libraries, plant DNA is isolated and fragments are generated, some of which will encode parts of the whole SCR protein. The DNA may be cleaved at specific sites using various restriction enzymes. Alternatively, one may use DNase in the presence of manganese to fragment the DNA, or the DNA can be physically sheared, as for example, by sonication. The DNA fragments can then be separated according to size by standard techniques, including, but not limited to, agarose and polyacrylamide gel electrophoresis, column chromatography and sucrose gradient centrifugation.

The genomic DNA or cDNA fragments can be inserted into suitable vectors, including, but not limited to, plasmids, cosmids, bacteriophages lambda or T_4, and yeast artificial chromosome (YAC) [See, for example, Sambrook et al., Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989); Glover, D.M(ed.), DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, U.K., Vols. I and II (1985)].

The *SCR* or *SCL* nucleotide probe, DNA or RNA, should be at least 17 nucleotides, preferably at least 26 nucleotides, and most preferably at least 50 nucleotides in length. The nucleotide probe is hybridized under moderate stringency conditions and washed either under moderate, or preferably under high stringency conditions. Clones in libraries with insert DNA having substantial homology to the *SCR* or *SCL* probe will hybridize to the probe. Hybridization of the nucleotide probe to genomic or cDNA libraries is carried out using methods known in the art. One of ordinary skill in the art will know that the appropriate hybridization and wash conditions depend on the length and base composition of the probe and that such conditions may be determined using

standard formulae. See, for example, Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Vol. 2, (1989) pp 11.45-11.57 and 15.55-15.57.

The identity of a cloned or amplified *SCR* gene sequence can be verified by comparing the amino acid sequences of its three open reading frames with the amino acid sequence of a *SCR* gene (*e.g.*, Arabidopsis SCR protein [SEQ ID No:2]). A *SCR* gene or coding sequence encodes a protein or polypeptide whose amino acid sequence is substantially similar to that of a SCR protein or polypeptide (*e.g.*, the amino acid sequence of any one of the SCR proteins and/or polypeptides shown in FIG. 5A, 5E, FIG. 8, FIG. 9, FIGS. 11A-B, FIGS. 15A-S, FIG. 17B, FIG. 18 and FIG. 25). The identity of the cloned or amplified *SCR* gene sequence may be further verified by examining its expression pattern, which should show highly localized expression in the embryo and/or root of the plant from which the *SCR* gene sequence was isolated.

Comparison of the amino acid sequences encoded by a cloned or amplified sequence may reveal that it does not contain the entire *SCR* gene or its promoter. In such a case, the cloned or amplified *SCR* gene sequence may be used as a probe to screen a genomic library for clones having inserts that overlap the cloned or amplified *SCR* gene sequence. A complete *SCR* gene and its promoter may be reconstructed by splicing the overlapping *SCR* gene sequences.

5.1.2. EXPRESSION OF *SCR* GENE PRODUCTS

SCR proteins, polypeptides and peptide fragments, mutated, truncated or deleted forms of SCR and/or SCR fusion proteins can be prepared for a variety of uses, including, but not limited to, the generation of antibodies, as reagents in assays, the identification of other cellular gene products involved in regulation of root development; etc.

SCR translational products include, but are not limited to, those proteins and polypeptides encoded by the

SCR gene sequences described in Section 5.1, above. The invention encompasses proteins that are functionally equivalent to the *SCR* gene products described in Section 5.1. Such a *SCR* gene product may contain one or more deletions, additions or substitutions of SCR amino acid residues within the amino acid sequence encoded by any one of the *SCR* gene sequences described, above, in Section 5.1, but which result in a silent change, thus producing a functionally equivalent *SCR* gene product. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity and/or the amphipathic nature of the residues involved.

For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine and glutamine; positively charged (basic) amino acids include arginine, lysine and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Functionally equivalent", as utilized herein, refers to a protein capable of exhibiting a substantially similar *in vivo* activity as the endogenous *SCR* gene products encoded by the *SCR* gene sequences described in Section 5.1, above. Alternatively, "functionally equivalent" may refer to peptides capable of regulating gene expression in a manner substantially similar to the way in which the corresponding portion of the endogenous *SCR* gene product would.

The invention also encompasses mutant SCR proteins and polypeptides that are not functionally equivalent to the gene products described in Section 5.1. Such a mutant *SCR* protein or polypeptide may contain one or more deletions, additions or substitutions of *SCR* amino acid residues within the amino acid sequence encoded by any one the SCR gene sequences described above in Section 5.1., and which result in loss of one or more functions of the SCR protein (e.g., recognition of a specific nucleic sequence, binding of a

transcription factor, etc.), thus producing a *SCR* gene product not functionally equivalent to the wild-type SCR protein.

While random mutations can be made to *SCR* DNA (using random mutagenesis techniques well known to those skilled in the art) and the resulting mutant SCRs tested for activity, site-directed mutations of the *SCR* gene and/or coding sequence can be engineered (using site-directed mutagenesis techniques well known to those skilled in the art) to generate mutant SCRs with increased function, (*e.g.*, resulting in improved root formation), or decreased function (*e.g.*, resulting in suboptimal root function). In particular, mutated SCR proteins in which any of the domains shown in FIGS. 13A-F are deleted or mutated are within the scope of the invention. Additionally, peptides corresponding to one or more domains of the SCR (*e.g.*, shown in FIGS. 13A-F), truncated or deleted SCRs, as well as fusion proteins in which the full length SCR, a SCR polypeptide or peptide fused to an unrelated protein are also within the scope of the invention and can be designed on the basis of the *SCR* nucleotide and SCR amino acid sequences disclosed in Section 5.1. above.

While the SCR polypeptides and peptides can be chemically synthesized (*e.g.*, see Creighton, 1983, Proteins: Structures and Molecular Principles, W.H. Freeman & Co., N.Y.) large polypeptides derived from SCR and the full length SCR may advantageously be produced by recombinant DNA technology using techniques well known to those skilled in the art for expressing nucleic acid sequences.

Methods which are well known to those skilled in the art can be used to construct expression vectors containing SCR protein coding sequences and appropriate transcriptional/translational control signals. These methods include, for example, *in vitro* recombinant DNA techniques, synthetic techniques and *in vivo* recombination/genetic recombination. See, for example, the techniques described in Sambrook et al., 1989, *supra*, and Ausubel et al., 1989,

supra. Alternatively, RNA capable of encoding SCR protein sequences may be chemically synthesized using, for example, synthesizers. See, for example, the techniques described in "Oligonucleotide Synthesis", 1984, Gait, M.J. ed., IRL Press, Oxford.

A variety of host-expression vector systems may be utilized to express the *SCR* gene products of the invention. Such host-expression systems represent vehicles by which the *SCR* gene products of interest may be produced and subsequently recovered and/or purified from the culture or plant (using purification methods well known to those skilled in the art), but also represent cells which may, when transformed or transfected with the appropriate nucleotide coding sequences, exhibit the SCR protein of the invention *in situ*. These include, but are not limited to, microorganisms such as bacteria (*e.g., E. coli, B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing SCR protein coding sequences; yeast (*e.g., Saccharomyces, Pichia*) transformed with recombinant yeast expression vectors containing the SCR protein coding sequences; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing the SCR protein coding sequences; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing *SCR* protein coding sequences; or mammalian cell systems (*e.g.*, COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (*e.g.*, metallothionein promoter) or from mammalian viruses (*e.g.*, the adenovirus late promoter; the vaccinia virus 7.5K promoter; the cytomegalovirus promoter/enhancer; etc.).

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the SCR protein being expressed. For example,

when a large quantity of such a protein is to be produced, for the generation of antibodies or to screen peptide libraries, for example, vectors which direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited, to the *E. coli* expression vector pUR278 (Ruther et al., 1983, EMBO J. 2:1791), in which the SCR coding sequence may be ligated individually into the vector in frame with the *lac Z* coding region so that a fusion protein is produced; pIN vectors (Inouye & Inouye, 1985, Nucleic Acids Res. 13:3101-3109; Van Heeke & Schuster, 1989, J. Biol. Chem. 264:5503-5509); and the like. pGEX vectors may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can be easily purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene protein can be released from the GST moiety.

In one such embodiment of a bacterial system, full length cDNA sequences are appended with in-frame Bam HI sites at the amino terminus and Eco RI sites at the carboxyl terminus using standard PCR methodologies (Innis et al., 1990, *supra*) and ligated into the pGEX-2TK vector (Pharmacia, Uppsala, Sweden). The resulting cDNA construct contains a kinase recognition site at the amino terminus for radioactive labelling and glutathione S-transferase sequences at the carboxyl terminus for affinity purification (Nilsson, et al., 1985, EMBO J. 4: 1075; Zabeau and Stanley, 1982, *EMBO J.* 1: 1217).

The recombinant constructs of the present invention may include a selectable marker for propagation of the construct. For example, a construct to be propagated in bacteria preferably contains an antibiotic resistance gene, such as one that confers resistance to kanamycin, tetracycline, streptomycin or chloramphenicol. Suitable

vectors for propagating the construct include plasmids, cosmids, bacteriophages or viruses, to name but a few.

In addition, the recombinant constructs may include plant-expressible, selectable or screenable marker genes for isolating, identifying or tracking plant cells transformed by these constructs. Selectable markers include, but are not limited to, genes that confer antibiotic resistance, (*e.g.*, resistance to kanamycin or hygromycin) or herbicide resistance (*e.g.*, resistance to sulfonylurea, phosphinothricin or glyphosate). Screenable markers include, but are not be limited to, genes encoding ß-glucuronidase (Jefferson, 1987, Plant Mol. Biol. Rep. 5:387-405), luciferase (Ow et al., 1986, Science 234:856-859) and B protein that regulates anthocyanin pigment production (Goff et al., 1990, EMBO J 9:2517-2522).

In embodiments of the present invention which utilize the *Agrobacterium tumefacien* system for transforming plants (see *infra*), the recombinant constructs may additionally comprise at least the right T-DNA border sequences flanking the DNA sequences to be transformed into the plant cell. Alternatively, the recombinant constructs may comprise the right and left T-DNA border sequences flanking the DNA sequence. The proper design and construction of such T-DNA based transformation vectors are well known to those skilled in the art.

5.1.3. ANTIBODIES TO *SCR* PROTEINS AND POLYPEPTIDES

Antibodies that specifically recognize one or more epitopes of SCR, or epitopes of conserved variants of SCR, or peptide fragments of the SCR are also encompassed by the invention. Such antibodies include, but are not limited to, polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, $F(ab')_2$ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies and epitope-binding fragments of any of the above.

For the production of antibodies, various host animals may be immunized by injection with the SCR protein, an SCR peptide (*e.g.*, one corresponding to a functional domain of the protein), a truncated SCR polypeptide (SCR in which one or more domains has been deleted), functional equivalents of the SCR protein or mutants of the SCR protein. Such SCR proteins, polypeptides, peptides or fusion proteins can be prepared and obtained as described in Section 5.1.2. *supra*. Host animals may include, but are not limited to, rabbits, mice and rats, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including, but not limited to, Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*. Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of the immunized animals.

Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, may be obtained by any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique of Kohler and Milstein, (Nature 256:495-497 [1975]; and U.S. Patent No. 4,376,110), the human B-cell hybridoma technique (Kosbor et al., 1983, Immunology Today 4:72; Cole et al., 1983, Proc. Natl. Acad. Sci. USA 80:2026-2030) and the EBV-hybridoma technique (Cole et al., 1985, Monoclonal Antibodies And Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). Such antibodies may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the mAb of this invention may be cultivated *in vitro* or *in vivo*. Production of high titers of mAbs *in vivo* makes this the presently preferred method of production.

In addition, techniques developed for the production of "chimeric antibodies" (Morrison et al., 1984, Proc. Natl. Acad. Sci., 81:6851-6855; Neuberger et al., 1984, Nature, 312:604-608; Takeda et al., 1985, Nature, 314:452-454) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region.

In addition, techniques have been developed for the production of humanized antibodies. (See, e.g., Queen, U.S. Patent No. 5,585,089.) An immunoglobulin light or heavy chain variable region consists of a "framework" region interrupted by three hypervariable regions, referred to as complementarity determining regions (CDRs). The extent of the framework region and CDRs have been precisely defined (see, "Sequences of Proteins of Immunological Interest", Kabat, E. et al., U.S. Department of Health and Human Services (1983)). Briefly, humanized antibodies are antibody molecules from non-human species having one or more CDRs from the non-human species and a framework region from a human immunoglobulin molecule.

Alternatively, techniques described for the production of single chain antibodies (U.S. Patent 4,946,778; Bird, 1988, Science 242:423-426; Huston et al., 1988, Proc. Natl. Acad. Sci. USA 85:5879-5883; and Ward et al., 1989, Nature 334:544-546) can be adapted to produce single chain antibodies against SCR proteins or polypeptides. Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.

Antibody fragments which recognize specific epitopes may be generated by known techniques. For example, such fragments include, but are not limited to: the $F(ab')_2$ fragments which can be produced by pepsin digestion of the

antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed (Huse et al., 1989, Science, 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Antibodies to a SCR protein and/or polypeptide can, in turn, be utilized to generate anti-idiotype antibodies that "mimic" SCR, using techniques well known to those skilled in the art. (See, *e.g.*, Greenspan & Bona, 1993, FASEB J 7(5):437-444; and Nissinoff, 1991, J. Immunol. 147(8):2429-2438).

5.1.4. *SCR* GENE OR GENE PRODUCTS AS MARKERS FOR QUANTITATIVE TRAIT LOCI

Any of the nucleotide sequences (including EST clone sequences) described in §§ 5.1 and 5.1.1. and/or listed in Tables 1 or 2, and/or polypeptides and proteins described in §§ 5.1.2. and/or listed in Tables 1 or 2, can be used as markers for quantitative trait loci in breeding programs for crop plants. To this end, the nucleic acid molecules, including, but not limited to, full length *SCR* coding sequences, and/or partial sequences (ESTs), can be used in hybridization and/or DNA amplification assays to identify the endogenous *SCR* genes, *scr* mutant alleles and/or *SCR* expression products in cultivars as compared to wild-type plants. They can be used also as markers for linkage analysis of quantitative trait loci. It is possible also that the *SCR* gene may encode a product responsible for a qualitative trait that is desirable in a crop breeding program. Alternatively, the *SCR* protein, peptides and/or antibodies can be used as reagents in immunoassays to detect expression of the *SCR* gene in cultivars and wild-type plants.

5.1.5. *SCR*-LIKE GENES

Scarecrow-like (*SCL*) genes are genes which show a high degree of similarity to the *SCR* gene. Tables 1 and 2 show a list of various *SCL* genes which were recently identified. Tables 1 and 2 also show each EST clone and/or genomic sequence corresponding with each of the *SCL* genes. The partial nucleotide sequence of various Arabidopsis EST's that encode members of the *SCL* gene family are shown in FIGS. 28A-AH.

Sequence analysis of the genes showed that a variable amino-terminal (N-terminal) and a highly conserved carboxyl-termini (C-termini) region exist throughout these putative gene products. The highly conserved region does not show significant similarity to members of any recognized gene family, indicating that these sequences likely define a novel gene family. Based on the high degree of similarity of the gene products to *SCR*, the genes corresponding to these ESTs were designated SCARECROW-LIKE (*SCL*). Recently, the importance of this gene family has been confirmed. Two components of the gibberellin signal transduction pathway, the gene products of the GIBBERELLIN-ACID INSENSITIVE (GAI) and the REPRESSOR OF GA1 (RGA) loci, have been shown to be members of this family (Peng et al., 1997, Genes & Dev. 11, 3194-3205; Silverstone et al., 1998, Plant Cell 10, 155-169). Thus, this family of gene products has been designated as the GRAS gene family, an acronym based on the designations of the known genes: GAI, RGA and SCR. An alignment of various GRAS gene products is shown at FIG. 29. As shown on the figure, the gene products have at least five recognizable motifs that are highly conserved. The absolutely conserved residues within the VHIID and SAW motifs are highlighted in bold, as are the hydrophobic residues of the leucine heptads, the P-F-Y-R-E residues of the PFYRE motif, and the two short sequences that define the end of the VHIID motif and the beginning of the PFYRE motif.

The GRAS family includes at present nineteen distinct members in Arabidopsis: fifteen SCLs, SCR, GAI, RGA, and RGAL (a GRAS sequence of unknown function with high similarity to GAI and RGA). The fact that the SCR, GAI, and RGA gene products have diverse roles in fundamental processes in plant biology (SCR in pattern formation and GAI/RGA in signal transduction) suggests that other members of this family may also play important roles in the physiology and development of higher plants. Intriguingly, the majority of the *SCL* genes are expressed predominantly in the root. FIG. 30 and Table 3. Furthermore, one of these (*SCL3*) has an expression pattern in the root that is similar to that of *SCR*. FIG. 31. In addition to root, many of the *SCL* genes are expressed in siliques and shoots. *See*, Table 3.

The *SCL* genes and gene products may be isolated and expressed with methods similar to those discussed for *SCR* genes at Sections 5.1.1. and 5.1.2., *supra*. Furthermore, antibodies to SCL proteins and polypeptides may be produced as was discussed in Section 5.1.3., *supra*. Finally, *SCL* genes and gene products may be used as markers for quantitative trait loci as was discussed at Section 5.1.4., *supra*.

	Length of EST (bp)	Estimated mRNA size (kb)	Expression of mRNA		
			Siliques	Shoots	Roots
SCL1	1359	1.5/1.7	+++++	+++++	+++++
SCL3	1231	1.8	++	++	+++
SCL5	1065	2.0	++	++	+++
SCL6	1279	2.4			+
SCL7	527	2.3	+	+	+
SCL8	1900	2.7	+	++	+++
SCL9	726	3.1			+
SCL11	760	2.1	+	++	+++
SCL13	1078	2.4	+	++	+++
SCL14	2635	3.2	++	++	++

Table 3

5.2. *SCR* PROMOTERS

According to the present invention, *SCR* promoters and functional portions thereof described herein refer to regions of the *SCR* gene which are capable of promoting tissue-specific expression in embryos, roots and shoots of an operably linked coding sequence in plants. The *SCR* promoter described herein refers to the regulatory elements of *SCR* genes, *i.e.*, regulatory regions of genes which are capable of selectively hybridizing to the nucleic acids described in Section 5.1, or regulatory sequences contained, for example, in the region between the translational start site of the Arabidopsis *SCR* gene and the HindIII site approximately 2.5 kb upstream of the site in plasmid pLIG1-3/SAC+Mob21SAC (see FIGS. 5A and 14) in hybridization assays, or which are homologous by sequence analysis (containing a span of 10 or more nucleotides in which at least 50 percent of the nucleotides are identical to the sequences presented herein). Homologous nucleotide sequences refer to nucleotide sequences including, but not limited to, *SCR* promoters in diverse plant species (*e.g.*, promoters of orthologs of Arabidopsis *SCR*) as well as genetically engineered derivatives of the promoters described herein.

Methods which could be used for the synthesis, isolation, molecular cloning, characterization and manipulation of *SCR* promoter sequences are well known to those skilled in the art. See, *e.g.*, the techniques described in Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd. ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989).

According to the present invention, *SCR* promoter sequences or portions thereof described herein may be obtained from appropriate plant or mammalian sources from cell lines or recombinant DNA constructs containing *SCR* promoter sequences, and/or by chemical synthetic methods. *SCR* promoter sequences can be obtained from genomic clones containing sequences 5' upstream of *SCR* coding sequences. Such 5' upstream clones may be obtained by screening genomic

libraries using SCR protein coding sequences, particularly those encoding SCR N-terminal sequences, from *SCR* gene clones obtained as described in Sections 5.1. and 5.2. Standard methods that may used in such screening include, for example, the method set forth in Benton & Davis, 1977, Science 196:180 for bacteriophage libraries; and Grunstein & Hogness, 1975, Proc. Nat. Acad. Sci. U.S.A. 72:3961-3965 for plasmid libraries.

The full extent and location of *SCR* promoters within such 5' upstream clones may be determined by the functional assay described below. In the event a 5' upstream clone does not contain the entire *SCR* promoter as determined by the functional assay, the insert DNA of the clone may be used to isolate genomic clones containing sequences further 5' upstream of the *SCR* coding sequences. Such further upstream sequences can be spliced on to existing 5' upstream sequences and the reconstructed 5' upstream region tested for functionality as a *SCR* promoter (*i.e.*, promoting tissue-specific expression in embryos and/or roots of an operably linked gene in plants). This process may be repeated until the complete *SCR* promoter is obtained.

The location of the *SCR* promoter within genomic sequences 5' upstream of the *SCR* gene isolated as described above may be determined using any method known in the art. For example, the 3' end of the promoter may be identified by locating the transcription initiation site, which may be determined by methods such as RNase protection (*e.g.*, Liang et al., 1989, J. Biol. Chem. 264:14486-14498), primer extension (*e.g.*, Weissenborn & Larson, 1992, J. Biol. Chem. 267:6122-6131) and/or reverse transcriptase/PCR. The location of the 3' end of the promoter may be confirmed by sequencing and computer analysis, examining for the canonical AGGA or TATA boxes of promoters that are typically 50-60 base pairs (bp) and 25-35 bp, respectively, 5' upstream of the transcription initiation site. The 5' end promoter may be defined by deleting sequences from the 5' end of the promoter containing fragment, constructing a transcriptional or

translational fusion of the resected fragment and a reporter gene and examining the expression characteristics of the chimeric gene in transgenic plants. Reporter genes that may be used to such ends include, but are not limited to, GUS, CAT, luciferase, β-galactosidase and C1 and R gene controlling anthocyanin production.

According to the present invention, a *SCR* promoter is one that confers to an operably linked gene in a transgenic plant tissue-specific expression in roots, root nodules, stems and/or embryos. A *SCR* promoter comprises the region between about -5,000 bp and +1 bp upstream of the transcription initiation site of a *SCR* gene. In a particular embodiment, the Arabidopsis *SCR* promoter comprises the region between positions -2.5 kb and +1 in the 5' upstream region of the Arabidopsis *SCR* gene (see FIGS. 5A and 14).

5.2.1. CIS-REGULATORY ELEMENTS OF *SCR* PROMOTERS

According to the present invention, the cis-regulatory elements within a *SCR* promoter may be identified using any method known in the art. For example, the location of cis-regulatory elements within an inducible promoter may be identified using methods such as DNase or chemical footprinting (*e.g.*, Meier et al., 1991, Plant Cell 3:309-315) or gel retardation (*e.g.*, Weissenborn & Larson, 1992, J. Biol. Chem. 267-6122-6131; Beato, 1989, Cell 56:335-344; Johnson et al., 1989, Ann. Rev. Biochem. 58:799-839). Additionally, resectioning experiments also may be employed to define the location of the cis-regulatory elements. For example, an inducible promoter-containing fragment may be resected from either the 5' or 3' end using restriction enzyme or exonuclease digests.

To determine the location of cis-regulatory elements within the sequence containing the inducible promoter, the 5' or 3' resected fragments, internal fragments to the inducible promoter containing sequence or inducible promoter fragments containing sequences identified by footprinting or gel retardation experiments may be fused to

the 5' end of a truncated plant promoter, and the activity of the chimeric promoter in transgenic plant examined. Useful truncated promoters to these ends comprise sequences starting at or about the transcription initiation site and extending to no more than 150 bp 5' upstream. These truncated promoters generally are inactive or are only minimally active. Examples of such truncated plant promoters may include, among others, a "minimal" CaMV 35S promoter whose 5' end terminates at position -46 bp with respect to the transcription initiation site (Skriver et al., Proc. Natl. Acad. Sci. USA 88:7266-7270); the truncated "-90 35S" promoter in the X-GUS-90 vector (Benfey & Chua, 1989, Science 244:174-181); a truncated "-101 nos" promoter derived from the nopaline synthase promoter (Aryan et al., 1991, Mol. Gen. Genet. 225:65-71); and the truncated maize Adh-1 promoter in pADcat 2 (Ellis et al., 1987, EMBO J. 6:11-16).

According to the present invention, a cis-regulatory element of a *SCR* promoter is a sequence that confers to a truncated promoter tissue-specific expression in embryos, stems, root nodules and/or roots.

5.2.2. *SCR* PROMOTER-DRIVEN EXPRESSION VECTORS

The properties of the nucleic acid sequences are varied as are the genetic structures of various potential host plant cells. In the preferred embodiments of the present invention, described herein, a number of features which an artisan may recognize as not being absolutely essential, but clearly advantageous are used. These include methods of isolation, synthesis or construction of gene constructs, the manipulation of the gene constructs to be introduced into plant cells, certain features of the gene constructs, and certain features of the vectors associated with the gene constructs.

Further, the gene constructs of the present invention may be encoded on DNA or RNA molecules. According to the present invention, it is preferred that the desired, stable genotypic change of the target plant be effected

through genomic integration of exogenously introduced nucleic acid construct(s), particularly recombinant DNA constructs. Nonetheless, according to the present invention, such genotypic changes also can be effected by the introduction of episomes (DNA or RNA) that can replicate autonomously and that are somatically and germinally stable. Where the introduced nucleic acid constructs comprise RNA, plant transformation or gene expression from such constructs may proceed through a DNA intermediate produced by reverse transcription.

The present invention provides for use of recombinant DNA constructs which contain tissue-specific and developmental-specific promoter fragments and functional portions thereof. As used herein, a functional portion of a *SCR* promoter is capable of functioning as a tissue-specific promoter in the embryo, stem, root nodule and/or root of a plant. The functionality of such sequences can be readily established by any method known in the art. Such methods include, for example, constructing expression vectors with such sequences and determining whether they confer tissue-specific expression in the embryo, stem, root nodule and/or root to an operably linked gene. In a particular embodiment, the invention provides for the use of the Arabidopsis *SCR* promoter contained in the sequences depicted in FIGS. 5A and 14 and the insert DNA of plasmid pGEX-2TK$^+$.

The *SCR* promoters of the invention may be used to direct the expression of any desired protein, or to direct the expression of a RNA product, including, but not limited to, an "antisense" RNA or ribozyme. Such recombinant constructs generally comprise a native *SCR* promoter or a recombinant *SCR* promoter derived therefrom, ligated to the nucleic acid sequence encoding a desired heterologous gene product.

A recombinant *SCR* promoter is used herein to refer to a promoter that comprises a functional portion of a native *SCR* promoter or a promoter that contains native promoter sequences that is modified by a regulatory element from a *SCR*

promoter. Alternatively, a recombinant inducible promoter derived from the *SCR* promoter may be a chimeric promoter, comprising a full-length or truncated plant promoter modified by the attachment of one or more *SCR* cis-regulatory elements.

The manner of chimeric promoter constructions may be any well known in the art. For examples of approaches that can be used in such constructions, see Section 5.1.2., above and Fluhr et al., 1986, Science 232:1106-1112; Ellis et al., 1987, EMBO J. 6:11-16; Strittmatter & Chua, 1987, Proc. Natl. Acad. Sci. USA 84:8986-8990; Poulsen & Chua, 1988, Mol. Gen. Genet. 214:16-23; Comai et al., 1991, Plant Mol. Biol. 15:373-381; Aryan et al., 1991, Mol. Gen. Genet. 225:65-71.

According to the present invention, where a *SCR* promoter or a recombinant *SCR* promoter is used to express a desired protein, the DNA construct is designed so that the protein coding sequence is ligated in phase with the translational initiation codon downstream of the promoter. Where the promoter fragment is missing 5' leader sequences, a DNA fragment encoding both the protein and its 5' RNA leader sequence is ligated immediately downstream of the transcription initiation site. Alternatively, an unrelated 5' RNA leader sequence may be used to bridge the promoter and the protein coding sequence. In such instances, the design should be such that the protein coding sequence is ligated in phase with the initiation codon present in the leader sequence, or ligated such that no initiation codon is interposed between the transcription initiation site and the first methionine codon of the protein.

Further, it may be desirable to include additional DNA sequences in the protein expression constructs. Examples of additional DNA sequences include, but are not limited to, those encoding: a 3' untranslated region; a transcription termination and polyadenylation signal; an intron; a signal peptide (which facilitates the secretion of the protein); or a transit peptide (which targets the protein to a particular cellular compartment such as the nucleus, chloroplast, mitochondria or vacuole).

5.3. PRODUCTION OF TRANSGENIC PLANTS AND PLANT CELLS

According to the present invention, a desirable plant or plant cell may be obtained by transforming a plant cell with the nucleic acid constructs described herein. In some instances, it may be desirable to engineer a plant or plant cell with several different gene constructs. Such engineering may be accomplished by transforming a plant or plant cell with all of the desired gene constructs simultaneously. Alternatively, the engineering may be carried out sequentially. That is, transforming with one gene construct, obtaining the desired transformant after selection and screening, transforming the transformant with a second gene construct, and so on.

In an embodiment of the present invention, *Agrobacterium* is employed to introduce the gene constructs into plants. Such transformations preferably use binary *Agrobacterium* T-DNA vectors (Bevan, 1984, Nuc. Acid Res. 12:8711-8721) and the co-cultivation procedure (Horsch et al., 1985, Science 227:1229-1231). Generally, the *Agrobacterium* transformation system is used to engineer dicotyledonous plants (Bevan et al., 1982, Ann. Rev. Genet. 16:357-384; Rogers et al., 1986, Methods Enzymol. 118:627-641). The *Agrobacterium* transformation system also may be used to transform, as well as transfer, DNA to monocotyledonous plants and plant cells (see Hernalsteen et al., 1984, EMBO J 3:3039-3041; Hooykass-Van Slogteren et al., 1984, Nature 311:763-764; Grimsley et al., 1987, Nature 325:1677-179; Boulton et al., 1989, Plant Mol. Biol. 12:31-40.; Gould et al., 1991, Plant Physiol. 95:426-434).

In other embodiments, various alternative methods for introducing recombinant nucleic acid constructs into plants and plant cells also may be utilized. These other methods are particularly useful where the target is a monocotyledonous plant or plant cell. Alternative gene transfer and transformation methods include, but are not limited to, protoplast transformation through calcium-,

polyethylene glycol (PEG), electroporation-mediated uptake of naked DNA (*see* Paszkowski et al., 1984, EMBO J 3:2717-2722, Potrykus et al., 1985, Mol. Gen. Genet. 199:169-177; Fromm et al., 1985, Proc. Natl. Acad. Sci. USA 82:5824-5828; Shimamoto, 1989, Nature 338:274-276) and electroporation of plant tissues (D'Halluin et al., 1992, Plant Cell 4:1495-1505). Additional methods for plant cell transformation include microinjection, silicon carbide mediated DNA uptake (Kaeppler et al., 1990, Plant Cell Reporter 9:415-418) and microprojectile bombardment (see Klein et al., 1988, Proc. Natl. Acad. Sci. USA 85:4305-4309; Gordon-Kamm et al., 1990, Plant Cell 2:603-618).

According to the present invention, a wide variety of plants may be engineered for the desired physiological and agronomic characteristics described herein using the nucleic acid constructs of the instant invention and the various transformation methods mentioned above. In preferred embodiments, target plants for engineering include, but are not limited to, crop plants such as maize, wheat, rice, soybean, tomato, tobacco, carrots, peanut, potato, sugar beets, sunflower, yam, Arabidopsis, rape seed and petunia; and trees such as spruce.

According to the present invention, desired plants and plant cells may be obtained by engineering the gene constructs described herein into a variety of plant cell types, including, but not limited to, protoplasts, tissue culture cells, tissue and organ explants, pollen, embryos as well as whole plants. In an embodiment of the present invention, the engineered plant material is selected or screened for transformants (*i.e.*, those that have incorporated or integrated the introduced gene construct(s)) following the approaches and methods described below. An isolated transformant may then be regenerated into a plant. Alternatively, the engineered plant material may be regenerated into a plant, or plantlet, before subjecting the derived plant, or plantlet, to selection or screening for the marker gene traits. Procedures for regenerating plants from

plant cells, tissues or organs, either before or after selecting or screening for marker gene(s), are well known to those skilled in the art.

A transformed plant cell, callus, tissue or plant may be identified and isolated by selecting or screening the engineered plant material for traits encoded by the marker genes present on the transforming DNA. For instance, selection may be performed by growing the engineered plant material on media containing inhibitory amounts of the antibiotic or herbicide to which the transforming marker gene construct confers resistance. Further, transformed plants and plant cells also may be identified by screening for the activities of any visible marker genes (*e.g.*, the ß-glucuronidase, luciferase, B or C1 genes) that may be present on the recombinant nucleic acid constructs of the present invention. Such selection and screening methodologies are well known to those skilled in the art.

Physical and biochemical methods also may be used to identify a plant or plant cell transformant containing the gene constructs of the present invention. These methods include, but are not limited to: 1) Southern analysis or PCR amplification for detecting and determining the structure of the recombinant DNA insert; 2) Northern blot, S-1 RNase protection, primer-extension or reverse transcriptase-PCR amplification for detecting and examining RNA transcripts of the gene constructs; 3) enzymatic assays for detecting enzyme or ribozyme activity, where such gene products are encoded by the gene construct; 4) protein gel electrophoresis, western blot techniques, immunoprecipitation, or enzyme-linked immunoassays, where the gene construct products are proteins; 5) biochemical measurements of compounds produced as a consequence of the expression of the introduced gene constructs. Additional techniques, such as *in situ* hybridization, enzyme staining, and immunostaining also may be used to detect the presence or expression of the recombinant construct in specific plant organs and tissues.

The methods for doing all of these assays are well known to those skilled in the art.

5.3.1. TRANSGENIC PLANTS THAT ECTOPICALLY EXPRESS *SCR*

In accordance with the present invention, a plant that expresses a recombinant *SCR* gene may be engineered by transforming a plant cell with a gene construct comprising a plant promoter operably associated with a sequence encoding a SCR protein or a fragment thereof. (Operably associated is used herein to mean that transcription controlled by the "associated" promoter would produce a functional messenger RNA, whose translation would produce the enzyme.) The plant promoter may be constitutive or inducible. Useful constitutive promoters include, but are not limited to, the CaMV 35S promoter, the T-DNA mannopine synthetase promoter and their various derivatives. Useful inducible promoters include, but are not limited to, the promoters of ribulose bisphosphate carboxylase (RUBISCO) genes, chlorophyll a/b binding protein (CAB) genes, heat shock genes, the defense responsive gene (*e.g.*, phenylalanine ammonia lyase genes), wound induced genes (*e.g.*, hydroxyproline rich cell wall protein genes), chemically-inducible genes (*e.g.*, nitrate reductase genes, gluconase genes, chitinase genes, PR-1 genes *etc.*), dark-inducible genes (*e.g.*, asparagine synthetase gene (Coruzzi and Tsai, U.S. Patent 5,256,558, October 26, 1993, Gene Encoding Plant Asparagine Synthetase)) and developmentally regulated genes (*e.g.*, Shoot Meristemless gene), to name just a few.

In yet another embodiment of the present invention, it may be advantageous to transform a plant with a gene construct operably linking a modified or artificial promoter to a sequence encoding a SCR protein or a fragment thereof. Typically, such promoters, constructed by recombining structural elements of different promoters, have unique expression patterns and/or levels not found in natural promoters. *See, e.g.*, Salina et al., 1992, Plant Cell

4:1485-1493, for examples of artificial promoters constructed from combining cis-regulatory elements with a promoter core.

In a preferred embodiment of the present invention, the associated promoter is a strong and root, root nodule, stem and/or embryo-specific plant promoter such that the SCR protein is overexpressed in the transgenic plant. Examples of root- and root nodules-specific promoters include, but are not limited to, the promoters of *SCR* genes, *SHR* genes, legehemoglobin genes, nodulin genes and root-specific glutamine synthetase genes (*See e.g.*, Tingey et al., 1987, EMBO J. 6:1-9; Edwards et al., 1990, Proc. Nat. Acad. Sci. USA 87:3459-3463).

In yet another preferred embodiment of the present invention, the overexpression of SCR protein in roots may be engineered by increasing the copy number of the *SCR* gene. One approach to producing such transgenic plants is to transform with nucleic acid constructs that contain multiple copies of the complete *SCR* gene (*i.e.*, with its own native *SCR* promoter). Another approach is to repeatedly transform successive generations of a plant line with one or more copies of the complete *SCR* gene. Yet another approach is to place a complete *SCR* gene in a nucleic acid construct containing an amplification-selectable marker (ASM) gene such as the glutamine synthetase or dihydrofolate reductase gene. Cells transformed with such constructs are subjected to culturing regimes that select cell lines with increased copies of complete *SCR* genes. *See, e.g.*, Donn et al., 1984, J. Mol. Appl. Genet. 2:549-562, for a selection protocol used to isolate a plant cell line containing amplified copies of the GS gene. Because the desired gene is closely linked to the ASM, cell lines that amplify the ASM gene are likely also to have amplified the *SCR* gene. Cell lines with amplified copies of the *SCR* gene can then be regenerated into transgenic plants.

5.3.2. TRANSGENIC PLANTS THAT SUPPRESS ENDOGENOUS *SCR* EXPRESSION

In accordance with the present invention, a desired plant may be engineered by suppressing *SCR* activity. In one embodiment, the suppression may be engineered by transforming a plant with a gene construct encoding an antisense RNA or ribozyme complementary to a segment, or the whole, of the *SCR* RNA transcript, including the mature target mRNA. In another embodiment, *SCR* gene suppression may be engineered by transforming a plant cell with a gene construct encoding a ribozyme that cleaves the *SCR* mRNA transcript. Alternatively, the plant can be engineered, *e.g.*, via targeted homologous recombination, to inactive or "knock-out" expression of the plant's endogenous *SCR*.

For all of the aforementioned suppression constructs, it is preferred that such gene constructs express specifically in the root, root nodule, stem and/or embryo tissues. Alternatively, it may be preferred to have the suppression constructs expressed constitutively. Thus, constitutive promoters, such as the nopaline and the CaMV 35S promoter, also may be used to express the suppression constructs. A most preferred promoter for these suppression constructs is a *SCR* or *SHR* promoter.

In accordance with the present invention, desired plants with suppressed target gene expression may be engineered also by transforming a plant cell with a co-suppression construct. A co-suppression construct comprises a functional promoter operatively associated with a complete or partial *SCR* gene sequence. It is preferred that the operatively associated promoter be a strong, constitutive promoter, such as the CaMV 35S promoter. Alternatively, the co-suppression construct promoter can be one that expresses with the same tissue and developmental specificity as the *SCR* gene.

According to the present invention, it is preferred that the co-suppression construct encodes an incomplete *SCR* mRNA, although a construct encoding a fully functional *SCR*

mRNA or enzyme also may be useful in effecting co-suppression.

In accordance with the present invention, desired plants with suppressed target gene expression also may be engineered by transforming a plant cell with a construct that can effect site-directed mutagenesis of the *SCR* gene. (*See, e.g.*, Offringa et al., 1990, EMBO J. 9:3077-84; and Kanevskii et al., 1990, Dokl. Akad. Nauk. SSSR 312:1505-1507 for discussions of nucleic constructs for effecting site-directed mutagenesis of target genes in plants.) It is preferred that such constructs effect suppression of the *SCR* gene by replacing the endogenous *SCR* gene sequence through homologous recombination with either none, or inactive SCR protein coding sequences.

5.3.3. TRANSGENIC PLANTS THAT EXPRESS A TRANSGENE CONTROLLED BY THE *SCR* PROMOTER

In accordance with the present invention, a desired plant may be engineered to express a gene of interest under the control of the *SCR* promoter. *SCR* promoters and functional portions thereof refer to regions of the nucleic acid sequence which are capable of promoting tissue-specific transcription of an operably linked gene of interest in the embryo, stem, root nodule and/or root of a plant. The *SCR* promoter described herein refers to the regulatory elements of *SCR* genes as described in Section 5.2.

Genes that may be beneficially expressed in the roots and/or root nodules of plants include genes involved in nitrogen fixation or cytokines or auxins, or genes which regulate growth, or growth of roots. In addition, genes encoding proteins that confer on plants herbicide, salt or pest resistance may be engineered for root specific expression. The nutritional value of root crops may be enhanced also through *SCR* promoter driven expression of nutritional proteins. Alternatively, therapeutically useful proteins may be expressed specifically in root crops.

Genes that may be beneficially expressed in the stems of plants include those involved in starch lignin or cellulose biosynthesis.

In accordance with the present invention, desired plants which express a heterologous gene of interest under the control of the *SCR* promoter may be engineered by transforming a plant cell with *SCR* promoter driven constructs using those techniques described in Section 5.2.2. and 5.3., *supra*.

5.3.4. SCREENING OF TRANSFORMED PLANTS FOR THOSE HAVING DESIRED ALTERED TRAITS

It will be recognized by those skilled in the art that in order to obtain transgenic plants having the desired engineered traits, screening of transformed plants (*i.e.*, those having an gene construct of the invention) having those traits may be required. For example, where the plants have been engineered for ectopic overexpression of a *SCR* gene, transformed plants are examined for those expressing the *SCR* gene at the desired level and in the desired tissues and developmental stages. Where the plants have been engineered for suppression of the *SCR* gene product, transformed plants are examined for those expressing the *SCR* gene product (*e.g.*, RNA or protein) at reduced levels in various tissues. The plants exhibiting the desired physiological changes, *e.g.*, ectopic *SCR* overexpression or *SCR* suppression, may then be subsequently screened for those plants that have the desired structural changes at the plant level (e.g., transgenic plants with overexpression or suppression of *SCR* gene having the desired altered root structure). The same principle applies to obtaining transgenic plants having tissue-specific expression of a heterologous gene in embryos and/or roots by the use of a *SCR* promoter driven expression construct.

Alternatively, the transformed plants may be directly screened for those exhibiting the desired structural and functional changes. In one embodiment, such screening may be for the size, length or pattern of the root of the

transformed plants. In another embodiment, the screening of the transformed plants may be for altered gravitropism or decreased susceptibility to lodging. In other embodiments, the screening of the transformed plants may be for improved agronomic characteristics (*e.g.*, faster growth, greater vegetative or reproductive yields or improved protein contents, etc.), as compared to unengineered progenitor plants, when cultivated under various growth conditions (*e.g.*, soils or media containing different amounts of nutrients and water content).

According to the present invention, plants engineered with *SCR* overexpression may exhibit improved vigorous growth characteristics when cultivated under conditions where large and thicker roots are advantageous. Plants engineered for *SCR* suppression may exhibit improved vigorous growth characteristics when cultivated under conditions where thinner roots are advantageous.

Engineered plants and plant lines possessing such improved agronomic characteristics may be identified by examining any of following parameters: 1) the rate of growth, measured in terms of rate of increase in fresh or dry weight; 2) vegetative yield of the mature plant, in terms of fresh or dry weight; 3) the seed or fruit yield; 4) the seed or fruit weight; 5) the total nitrogen content of the plant; 6) the total nitrogen content of the fruit or seed; 7) the free amino acid content of the plant; 8) the free amino acid content of the fruit or seed; 9) the total protein content of the plant; and 10) the total protein content of the fruit or seed. The procedures and methods for examining these parameters are well known to those skilled in the art.

According to the present invention, a desired plant is one that exhibits improvement over the control plant (*i.e.*, progenitor plant) in one or more of the aforementioned parameters. In an embodiment, a desired plant is one that shows at least 5% increase over the control plant in at least one parameter. In a preferred embodiment, a desired plant is one that shows at least 20% increase over the control plant

in at least one parameter. Most preferred is a plant that shows at least 50% increase in at least one parameter.

6. EXAMPLE 1: ARABIDOPSIS *SCR* GENE

This example describes the cloning and structure of the Arabidopsis *SCR* gene and its expression. The deduced amino acid sequence of the Arabidopsis *SCR* gene product contains a number of potential functional domains similar to those found in transcription factors. Closely related sequences have been found in both dicots and monocots indicating that Arabidopsis SCR is a member of a new protein family. The expression pattern of the *SCR* gene was characterized by means of *in situ* hybridization and by an enhancer trap insertion upstream of the *SCR* gene (described in more detail in Section 7). The expression pattern is consistent with a key role for Arabidopsis *SCR* in regulating the asymmetric division of the cortex/endodermis initial which is essential for generating the radial organization of the root.

6.1. MATERIALS AND METHODS

6.1.1. PLANT CULTURE

Arabidopsis ecotypes Wassilewskija (Ws), Columbia (Col), and Landsberg *erecta* (Ler) were obtained from Lehle. Arabidopsis seeds were surface sterilized and grown as described previously (Benfey et al., 1993, Development 119:57-70). Generation of the enhancer trap lines is described in Section 7.

6.1.2. GENETIC ANALYSIS

For the *scr-1* allele, co-segregation of the mutant phenotype and kanamycin resistance conferred by the inserted T-DNA was determined as described previously (Aeschbacher et al., 1995, Genes & Development 9:330-340). Because kanamycin affects root growth, 1557 seeds from heterozygous lines were germinated on non-selective media, scored for the appearance of the mutant phenotype, and subsequently transferred to

selective media. All (284) phenotypically mutant seedlings showed resistance to the antibiotic, whereas 834 of 1273 phenotypically wild-type seedlings showed resistance to kanamycin, respectively. Phenotypically wild type plants (83) were also transferred to soil and allowed to set seeds. The progeny of these plants were plated on selective and non-selective media, and scored for the co-segregation of the mutant phenotype and antibiotic resistance. A majority (48) of the plants segregated for the mutant phenotype and for kanamycin resistance, whereas 35 were wild-type and sensitive to kanamycin. Due to a mis-identified cross, *scr-2* was originally thought to be non-allelic and was named *pinocchio* (Scheres et al., 1995, Development 121:53-62). Subsequent mapping results placed it in an identical chromosomal location as *scr-1*. The original *scr-2* line contained at least two T-DNA inserts. Co-segregation analysis revealed a lack of linkage between the antibiotic resistance marker carried by the T-DNA and the mutant phenotype. Antibiotic sensitive lines were identified that segregated for mutants. These lines were crossed to *scr-1*. All F1 antibiotic resistant progeny exhibited a mutant phenotype. All F2 progeny (from independent lines) were mutant, and there was a 3:1 segregation for antibiotic resistance indicating that the two mutations were allelic. Antibiotic sensitive lines of *scr-2* were found to contain a rearranged T-DNA insert as determined by Southern blots and PCR using T-DNA specific probes and primers, respectively. The presence of this T-DNA in the *SCR* gene was confirmed by Southern blots using *SCR* probes. A combination of T-DNA and *SCR* specific primers was used to amplify T-DNA/*SCR* junctions. The PCR fragments were cloned using the TA cloning kit (Invitrogen) and sequenced. The insertion points were determined for both 5' and 3' T-DNA/*SCR* junctions.

6.1.3. MAPPING

Mutant plants of *scr-2* (WS background) were crossed to Col WT. DNA from mutant F2 individual plants were analyzed for co-segregation with microsatellite (Bell & Ecker, 1994, Genomics 18:137-144) and CAPS markers (Konieczny & Ausubel, 1993, Plant J. 4:403-410). The closest linkage was found to two CAPS markers located at the bottom of chromosome III. Only one out of 238 mutant chromosomes was recombinant for the BGL1 marker (Konieczny & Ausubel, 1993, Plant J. 4:403-410) and one out of 210 chromosomes was recombinant for the cdc2b marker.

A RFLP for the *SCR* gene was identified between Col and Ler ecotypes with Xho I endonuclease. Genomic DNAs from independent R1 lines (Jarvis et al., 1994, Plant Mol. Biol. 24:685-687) were digested with Xho I and blots were hybridized to *SCR*. Using the segregation data obtained for 25 R1 lines, the *SCR* gene was mapped relative to molecular markers by CLUSTER. The *SCR* gene was assigned to the bottom of chromosome III closest to BGL1.

6.1.4. PHENOTYPIC ANALYSIS

Morphological characterization of the mutant roots was performed as follows: 7 to 14 days post-germination, phenotypically mutant seedlings were fixed in 4.0% formaldehyde in PIPES buffer pH 7.2. After fixation, the samples were dehydrated in ethanol followed by infiltration with Historesin (Jung-Leica, Heidelberg, Germany). Plastic sections were mounted on superfrost slides (Fisher). The sections were either stained with 0.05% toluidine blue and photographed using Kodak 160T film, or used for Casparian strip detection or antibody staining.

Casparian strip detection was performed as described previously (Scheres et al., 1995, Development 121:53-62), with the following modifications. Plastic sections were used and the counterstaining was done in 0.1% aniline blue for 5 to 15 min. The sections were visualized with a Leitz fluorescent microscope with a FITC filter.

Pictures were taken using a Leitz camera attached to the microscope and Kodak HC400 film. Slides were digitized with a Nikon slide scanner and manipulated in Adobe Photoshop.

For antibody staining, sections were blocked for 2 hours at room temperature in 1% BSA in PBS containing 0.1% Tween 20 (PBT). Samples were incubated with primary antibodies at 4° C in 1% BSA in PBT overnight, and then washed 3 times 5 minutes each with PBT. Samples were incubated for two hours with biotinylated secondary antibodies (Vector Laboratories) in PBT, and washed as above. Samples were incubated with Texas Red conjugated avidin D for 2 hours at room temperature, washed as before, and mounted in Citifluor. Immunofluorescence was observed with a fluorescent microscope equipped with a Rhodamine filter. Staining with the CCRC antibodies was performed as described previously (Freshour et al., 1996, Plant Physiol. 110:1413-1429).

6.1.5. MOLECULAR TECHNIQUES

Genomic DNA preparation was performed using the Elu-Quik kit (Schleicher & Schuell) protocol. Radioactive and non-radioactive DNA probes were labeled with either random primed labeling or PCR-mediated synthesis according to the Genius kit manual (Boehringer Mannheim). *E. coli* and *Agrobacterium tumefaciens* cells were transformed using a BIO-RAD gene pulser. Plasmid DNA was purified using the alkaline lysis method (Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, New York:Cold Spring Harbor Laboratory, 1982).

A probe made from a rescued fragment of 1.2 kb was used to screen a wild-type genomic library made from WS plants. One genomic clone containing an insert of approximately 23 kb was isolated. A 3.0 kb Sac I fragment from the genomic clone, which hybridized to the 1.2 kb probe, was subcloned and sequenced (FIG. 5A). Comparison of the nucleotide sequence between the genomic clone and the rescued plasmid revealed the site of the T-DNA insertion.

Approximately 600,000 plaques from a cDNA library, obtained from inflorescences and siliques (Col ecotype), and therefore enriched in embryos, were screened with the 1.2 kb probe. Four cDNA clones were isolated. The dideoxy sequencing method was performed using the Sequenase kit (United States Biochemical Corp.). Sequence-specific internal primers were synthesized and used to sequence the Sac I genomic as well the cDNA clones. Total RNA from plant tissues was obtained using phenol/chloroform extractions as described in Berry et al., 1985, Mol. Cell. Biol. 5:2238-2246 with minor modifications. Northern hybridization and detection were performed according to the Genius kit manual (Boehringer Mannheim).

To identify the site of insertion of the enhancer-trap T-DNA, genomic DNA from ET199 homozygous plants was amplified using primers specific for the T-DNA left border and the *SCR* gene. An approximately 2.0 kb fragment was amplified. This fragment was sequenced and the site of insertion was found to be approximately 1 kb from the ATG start codon.

6.1.6. *IN SITU* HYBRIDIZATION

Antisense and sense *SCR* riboprobes were labeled with digoxigenin-11-UTP (Boehringer Mannheim) using T7 polymerase following the manufacturer's protocol. Probes contained a 1.1 kb 3' portion of the cDNA. Probe purification, hydrolysis and quantification were performed as described in the Boehringer Mannheim Genius System user's guide.

Tissue samples were fixed in 4 % formaldehyde overnight at 4°C and rinsed two times in PBS (Jackson et al., 1991, Pl. Cell 3:115-125). They were subsequently pre-embedded in 1 % agarose in PBS. The fixed tissue was dehydrated in ethanol, cleared in Hemo-De (Fisher Scientific, Pittsburgh, PA) and embedded in ParaplastPlus (Fisher Scientific). Tissue sections (10μm thick) were mounted on SuperfrostPlus slides (Fisher Scientific). Section

pretreatment and hybridization were performed according to Lincoln et al., 1994, Plant Cell 6:1859-1876 except that proteinase K was used at 30 mg/ml and a two hour prehybridization step was included. A probe concentration of 50 ng/ml/kb was used in the hybridization.

Slides were washed and the immunological detection was performed according to Coen et al., 1990, Cell 63:1311-1322 with the following modifications. Slides were first washed 5 hours in 5xSSC, 50% formamide. After RNase treatment, slides were rinsed three times (20 min each) in buffer (0.5 M NaCl, 10 mM Tris-HCl pH 8.0, 5.0 mM EDTA). In the immunological detection, antibody was diluted 1:1000, levamisole (240 ng/ml) was included in the detection buffer, and after stopping the reaction in 10 mM Tris, 1 mM EDTA, sections were mounted directly to Aqua-Poly/Mount (Polysciences, Warrington, PA).

6.2. RESULTS

6.2.1. CHARACTERIZATION OF THE *SCR* PHENOTYPE

The *scarecrow* mutant *scr-1* was isolated in a screen of T-DNA transformed Arabidopsis lines (Feldmann, K.A., 1991, Plant J. 1:71-82), as a seedling with greatly reduced root length compared to wild-type (Scheres et al., 1995, Development 121:53-62). A second mutant *scr-2* with a similar phenotype was subsequently identified among T-DNA transformed lines. Analysis of co-segregation between the mutant phenotype and antibiotic resistance carried by the T-DNA indicated tight linkage for *scr-1* and no linkage for *scr-2* (see Experimental Procedures). An antibiotic sensitive line of *scr-2* was isolated and crossed with *scr-1*. The F2 progeny of this cross were all mutant and segregated 3:1 for antibiotic resistance confirming allelism (see Materials & Methods). The principal phenotypic difference between the two alleles was that *scr-1* root growth was more retarded than that of *scr-2*, suggesting that it is the stronger allele (FIG. 2A). For both alleles, the aerial organs appeared similar to wild-type and the flowers were fertile (FIGS. 2A

and 2B). The progeny of backcrosses of *scr-1* or *scr-2* to wild-type plants segregated 3:1 for the root phenotype for both alleles, indicating that each mutation is monogenic and recessive.

Analysis of transverse sections through the primary root of seedlings revealed only a single cell layer between the epidermis and the pericycle (FIG. 2C) instead of the normal radial organization consisting of cortex and endodermis (FIG. 2D). This radial organization defect was not limited to the primary root, but also was present in secondary roots (FIG. 2E) and in roots regenerated from calli (FIG. 2F). Occasionally, defects were observed in the number of cells in the remaining cell layer (more than the invariant eight (8) found in wild-type). Abnormal placement or numbers of epidermal cells also were observed (see FIG. 2E). These abnormalities were more frequently observed in *scr-1* than in *scr-2*. Nevertheless, organization of the mutant root closely resembles that of wild-type except for the consistent reduction in the number of cell layers. Because the endodermis and cortex are normally generated by an asymmetric division of the cortex/endodermal initial, this indicates that the primary defect in *scr* is disruption of this asymmetric division.

It has been shown that the radial organization defect in *scr-1* first appears in the developing embryo at the early torpedo stage and manifests itself as a failure of the embryonic ground tissue to undergo the asymmetric division into cortex and endodermis (Scheres et al., 1995, Development 121:53-62). This defect extends the length of the embryonic axis which encompasses the embryonic root and hypocotyl. Other embryonic tissues appear similar to wild-type (Scheres et al., 1995, Development 121:53-62). In seedling hypocotyls of the *scarecrow* phenotype, two cell layers instead of the normal three layers (two cortex and one endodermis) between epidermis and stele were found. This would be the expected result of the lack of the division of the embryonic ground tissue. Similar results were obtained for *scr-2*. Hence,

this mutant identifies a gene involved in the asymmetric division that produces cortex and endodermis from ground tissue in the embryonic root and hypocotyl and from the cortex/endodermal initials in primary and secondary roots.

6.2.2. CHARACTERIZATION OF CELL IDENTITY IN *SCR* ROOTS

To understand the role of the Arabidopsis *SCR* gene in regulating this asymmetric division, it was necessary to determine the identity of the mutant cell layer. Tissue-specific markers were used to distinguish between several possibilities. The cell layer could have differentiated attributes of either cortex or endodermis. Alternatively, it could have an undifferentiated, initial-cell identity or it could have a chimeric identity with differentiated attributes of both endodermis and cortex in the same cell.

Transverse sections of *scr-1* and *scr-2* roots were assayed for the presence of tissue-specific markers. The casparian strip, a deposition of suberin between radial cell walls, is specific to endodermal cells and is believed to act as a barrier to the entry of solutes into the vasculature (Esau, K. Anatomy of Seed Plants, New York: John Wiley & Sons, 1977, Ed. 2, pp. 1-550). Histochemical staining revealed the presence of a casparian strip in the mutant cell layer (FIG. 3A, compare to wild-type, FIG. 3B). It is noted that in the vascular cylinder, this histochemical stain also reveals the presence of lignin, indicating the presence of differentiated xylem cells in mutant (FIG. 3A) and wild-type (FIG. 3B). Another marker of the differentiated endodermis is the arabinogalactan epitope recognized by the monoclonal antibody, JIM13 (Knox et al., 1990, Planta 181:512-521). The mutant cell layer showed staining with this antibody (FIG. 3C, compare with wild-type, FIG. 3B). As a positive control, the JIM7 antibody that recognizes pectin epitopes in all cell walls was used (FIGS. 3E and 3F). These results indicate that the cell layer between the epidermis and the pericycle has differentiated attributes of the endodermis.

As a marker for the cortex, the CCRC-M2 monoclonal antibody was used. This antibody recognizes a cell wall oligosaccharide epitope, found only on differentiated cortex and epidermis cells. In sections from the differentiation zone of *scr-1* and *scr-2*, both cortex and epidermal cells showed staining (FIG. 4A and 4B) that was similar to that of wild-type (FIG. 4C). In *scr-1*, staining of both cell types was apparent, but staining of cortex was somewhat weaker than wild-type. The positive control used the CCRC-M1 monoclonal antibody which recognizes an oligosaccharide epitope found on all cells (FIGS. 4D-F).

With the CCRC-M2 antibody, an interesting difference was observed between the staining pattern of the mutants as compared to wild-type. The appearance of this epitope correlates with differentiation in these two cell types. Normally, in sections close to the root tip, there is no staining. In sections higher up in the root, atrichoblasts (epidermal cells that do not make root hairs) stain. In sections from more mature root tissue, all epidermal cells as well as cortex cells stain for this epitope. In both *scr-1* and *scr-2*, sections could be found in which all epidermal cells stained while there was little detectable staining of cortex cells. Although not precisely identical to the wild-type staining pattern, the fact that the mutant cell layer clearly stains for this cortex marker indicates that there are cortex differentiated attributes expressed in these cells.

Taken together, these results indicate that the mutant cell layer has differentiated attributes of both the endodermis and cortex. The possibility that there has been a simple deletion of a cell type, or that the resulting cell type remains in an undifferentiated initial-like stage can be ruled out. This result is consistent with a role for the *SCR* gene in regulating this asymmetric division rather than a role in directing cell specification.

6.2.3. MOLECULAR CLONING OF THE *SCR* GENE

To further elucidate the function of the Arabidopsis *SCR* gene, the inserted T-DNA sequences were used to clone the gene. Plant DNA flanking the insertion site was obtained from *scr-1* by plasmid rescue and used to isolate the corresponding wild-type genomic DNA. Several cDNA clones were isolated from a library made from silique tissue. Comparison of the sequence of the longest cDNA and the corresponding genomic region revealed an open reading frame (ORF) interrupted by a single small intron. (FIG. 5A). A potential TATA box and polyadenylation signal that matched the consensus sequences for plant genes were also identified (Joshi, C.P., 1987, Nucl. Acids Res. 15:6643-6653); Heidecker & Messing, 1986, Ann. Rev. Plant Physiol. 37:439-466); Mogen et al., 1990, Plant Cell 2:1261-1272).

Comparison of the nucleotide sequence between the genomic clone and the rescued plasmid placed the site of the T-DNA insertion in *scr-1* at codon 470 (FIGS. 5A and 5B). For *scr-2*, although no linkage was found between the mutant phenotype and antibiotic resistance, DNA blot and PCR analysis of antibiotic sensitive lines revealed the presence of T-DNA sequences that co-segregated with the mutant phenotype. The insertion position in *scr-2* was determined by cloning and sequencing the PCR products amplified from its genomic DNA using a combination of T-DNA and *SCR* specific primers at both sides of the insertion (FIG. 5B). In *scr-2*, the T-DNA insertion point is at codon 605 (FIG. 5A and 5B).

To verify linkage between the cloned gene and the mutant phenotype, we identified the chromosomal location of both the *scr* locus and the *SCR* gene. To map the *scr* locus, molecular markers were used on F2 progeny of crosses between *scr-2* (ecotype Wassilewskija, Ws) and Colombia (Col) WT. These placed the *scr* locus at the bottom of chromosome III, approximately 0.5 cM away from each of the two closest markers available, cdc2b and BGL1 (Konieczny and Ausubel, 1993, Plant J. 4:403-410). To map the *SCR* gene, we identified a polymorphism between Col and Landsberg (Ler)

ecotypes using the *SCR* probe b (FIG. 5B). Southern analysis of 25 recombinant inbred lines (Jarvis et al., 1994, Plant Mol. Biol. 24:685-687) mapped the cloned gene to the same location as the *SCR* locus on chromosome III.

The determination of the molecular defects in two independent alleles and the co-localization of the cloned gene and the mutant locus confirms that we have identified the *SCR* gene.

6.2.4. THE *SCR* GENE HAS MOTIFS THAT INDICATE IT IS A TRANSCRIPTION FACTOR

The Arabidopsis *SCR* gene product is a 653 amino acid polypeptide that contains several domains (FIG. 5B). The amino-terminus has homopolymeric stretches of glutamine, serine, threonine and proline residues, which account for 44% of the first 267 residues. Domains rich in these residues have been shown to activate transcription and may serve such a role in SCR (Johnson et al., 1993, J. Nutr. Biochem 4:386-398). A charged region between residues 265 and 283 has similarity to the basic domain of the bZIP family of transcriptional regulatory proteins (FIG. 5C) (Hurst, H.C., 1994, Protein Profile 1:123-168). The basic domains from several bZIP proteins have been shown to act as nuclear localization signals (Varagona et al., 1992, Plant Cell 4:1213-1227), and this region in SCR may act similarly. This charged region is followed by a leucine heptad repeat (residues 291-322). A second leucine heptad repeat is found toward the carboxy-terminus (residues 436 to 473). As leucine heptad repeats have been demonstrated to mediate protein-protein interactions in other proteins (Hurst, H.C., 1994, Protein Profile 1:123-168), the existence of these motifs suggests that SCR may function as a dimer or a multimer. The second leucine heptad repeat is followed by a small region rich in acidic residues, also present in a number of defined transcriptional activation domains (Johnson et al., 1993, J. Nutr Biochem 4:386-398). While each of these domains has been found within proteins that do not act

as transcriptional regulators, the fact that all of them are found within the deduced SCR protein sequence indicates that SCR is a transcriptional regulatory protein.

6.2.5. SCR IS A MEMBER OF A NOVEL PROTEIN FAMILY

The Arabidopsis SCR protein sequence was compared with the sequences in the available databases. Eleven expressed sequence tags (ESTs), nine from Arabidopsis, one from rice and one from maize, showed significant similarity to residues 394 to 435 of the SCR sequence, a region immediately amino-terminal to the second leucine heptad repeat (FIGS. 15K-L). This region is designated the VHIID domain. Subsequent analysis of these EST sequences has revealed that the sequence similarity extends beyond this region; in fact, the similarity extends throughout the entire known gene products. The combination and order of the motifs found in these sequences do not show significant similarity to the general structures of other established regulatory protein families (*i.e.*, bZIP, zinc finger, MADS-domain and homeodomain), indicating that the SCR proteins comprise a novel family.

6.2.6. *SCR* IS EXPRESSED IN THE CORTEX/ENDODERMAL INITIALS AND IN THE ENDODERMIS

RNA blot analysis revealed expression of *SCR* in Arabidopsis siliques, leaves and roots of wild-type plants (FIG. 6A). No hybridization was detected to RNA from *scr-1* plants (FIG. 6B, lane 2). This indicates that *scr-1* has a reduced level of RNA expression and may represent the null phenotype. Hybridization to RNA species larger than the normal size were detected in *scr*-2. This indicates that abnormal *SCR* transcripts are made in this allele, suggesting that functional but possibly altered proteins may be produced.

To determine if expression was localized to any particular cell type, RNA *in situ* hybridization was performed on sections of root tissue. In mature roots, expression was

localized primarily to the endodermis (FIGS. 7A and 7B). Expression appeared to start very close to, or within, the cortex/endodermal initials and continue up the endodermal cell file as far as the section extended. Expression was detected also in late-torpedo stage embryos in the endodermis throughout the embryonic axis (FIG. 7C). Sense strand controls showed only background hybridization (FIG. 7D).

To determine whether the localization of *SCR* RNA was regulated at the transcriptional or post-transcriptional level, enhancer trap (ET) lines were prepared and examined in which the β-glucuronidase (uid-A or GUS) coding sequence with a minimal promoter was expressed in the root endodermis. (See Section 7, *infra*). Restriction fragment length polymorphisms were observed when DNA from one of these lines, ET199 and wild-type were probed with *SCR*. PCR and sequence analysis confirmed that the enhancer-trap construct had inserted approximately 1 kb upstream of the *SCR* start site and in the same orientation as that of *SCR* transcription.

In mature roots, expression in ET199 whole mounts showed a similar pattern to that of the *in situ* hybridizations, with the strongest staining present in endodermal cells (FIG. 7E). Transverse sections indicated that expression was primarily in endodermal cells in the elongation zone (FIG. 7F). Longitudinal sections through the meristematic zone revealed that expression could be detected in the cortex/endodermal initial (FIG. 7G). Of particular interest was the restriction of expression to the endodermal daughter cell after the periclinal division (FIG. 7G). This indicated that the expression pattern observed in the *in situ* analysis was not due to post-transcriptional partitioning of *SCR* RNA. Rather, it suggests that after the periclinal division of the cortex/endodermis initial, only one of the two cells is able to transcribe *SCR* RNA.

6.3. DISCUSSION

6.3.1. THE *SCR* GENE REGULATES AN ASYMMETRIC DIVISION REQUIRED FOR ROOT RADIAL ORGANIZATION

The formation of the cortex and endodermal layers in the Arabidopsis root requires two asymmetric divisions. In the first, an anticlinal division of the cortex/endodermal initial generates two cells with different developmental potentials. One will continue to function as an initial, while the other undergoes a periclinal division to generate the first cells in the endodermal and cortex cell files. This second asymmetric division is eliminated in the *scarecrow* mutant, resulting in a single cell layer instead of two. The *scr* mutation appears to have little effect on any other cell divisions in the root indicating that it is involved in regulating a single asymmetric division in this organ. Several other mutations have been characterized that appear to affect specific cell division pathways in Arabidopsis. These include *knolle* (*kn*), in which formation of the epidermis is impaired (Lukowitz et al., 1996, Cell 84:61-71); *wooden leg* (*wol*), in which vascular cell division is defective (Scheres et al., 1995, Development 121:53-62) and *fass* (*fs*), in which there are supernumerary cortex and vascular cells (Scheres et al., 1995, Development 121:53-62); Torres Ruiz & Jurgens, 1994, Development 120:2967-2978). Only in the case of *scr* and *short-root* (*shr*) mutants has it been shown that the defect is in a specific asymmetric division.

Mutational analyses in several organisms have revealed that the genes that regulate asymmetric divisions can be specific to a single type of division or can affect divisions that are not clonally related (Horvitz & Herskowitz, 1992, Cell 68:237-255). In most cases, these mutations result in the formation of two identical daughter cells with similar developmental potentials (Horvitz & Herskowitz, 1992, Cell 68:237-255). Both resulting cells have the identity of one or the other of the normal daughter

cells, an example of which is the swi mutation in *S. cerevisiae* (Nasmyth et al., 1987, Cell 48:579-587). However, there are also examples of mutations that result in the formation of chimeric cell types such as the *ham-1* mutation in *C. elegans* (Desai et al., 1988, Nature 336:638-646).

6.3.2. *SCR* INVOLVEMENT IN CELL SPECIFICATION OR CELL DIVISION

Genes that regulate asymmetric cell divisions can be divided into those that specify the differentiated fates of the daughter cells and those that function to effect the division of the mother cell (Horvitz & Herskowitz, 1992, Cell, 68:237-255). The aberrant cell layer formed in the *scr* mutant has differentiated features of both endodermal and cortex cells. Thus, *scr* is in the rare class of asymmetric division mutants in which a chimeric cell type is created. The ability to express differentiated characteristics of cortex and endodermal cells implies that the differentiation pathways for both of these cell types are intact and do not require the functional *SCR* gene. This indicates that *SCR* is involved primarily in regulating a specific cell division, and that the correct occurrence of this division can be unlinked from cell specification. This is in contrast to the *shr* mutant, in which the periclinal division of the cortex/endodermal initial also fails to occur and the resulting cell lacks endodermal markers (Benfey et al., 1993, Development 119:57-70) and has cortex attributes. A genetic analysis was used to address the function of *SHR* and *SCR* in the asymmetric division of the cortex/endodermal initial. Placing mutants of each of these genes in a *fs* mutant background answered whether the supernumerary cell divisions characteristic of *fs* were sufficient to restore normal cell identities (Scheres et al., 1995, Development 121:53-62). In the *shr,fs* double mutant, there were additional cell layers but no endodermal, indicating that the *SHR* gene has a role in specifying cell identity. In the *scr,fs* double mutant, no

alteration in cell identity was observed as compared to *fs* (Scheres et al., 1995, Development 121:53-62). Taken together with the cell marker analysis presented herein, these results are consistent with a role for *SCR* in generating the division of the mother cell while the *SHR* gene may be involved in specifying the fate of the endodermal daughter.

6.3.3. A ROLE FOR *SCR* IN EMBRYONIC DEVELOPMENT

At least one additional cell division appears to be affected in the *scr* mutant. During embryonic development, the ground tissue does not divide to form the endodermal and cortex layers of the embryonic root and hypocotyl. As shown herein, expression of *SCR* was detected in the endodermal tissue throughout the embryonic axis shortly after this division occurs. Thus, *SCR* may play a direct role in regulating both this division and the division of the cortex/endodermal initial in the root apical meristem. Alternatively, the radial organization established in the embryo may somehow act as a template that directs the division of the cortex/endodermal initial, thus perpetuating the pattern. This is consistent with the finding in the *scr* mutant that the aberrant pattern established in the embryo is perpetuated in the primary root. It also is consistent with a recent study in which the daughter cells of the cortex/endodermal initial were laser ablated (van den Berg et al., 1995, Nature 378:62-65). When a single daughter cell was ablated, it was replaced by a cell that followed the normal asymmetric division pattern. When three adjacent daughter cells were ablated, the central initial divided anticlinally but failed to perform the periclinal division (van den Berg et al., 1995, Nature 378:62-65). This provided evidence that information from mature cells is required for the correct division pattern of cortex/endodermal initials suggesting a "top down" transfer of information. However, the absence of a cell layer in lateral roots and callus-derived roots of the *scr* mutant suggests that embryo events

are not unique in their ability to establish radial organization. Rather, these observations implicate *SCR* in regulating both embryonic and post-embryonic root radial organization.

6.3.4. TISSUE-SPECIFIC EXPRESSION OF *SCR* IS REGULATED AT THE TRANSCRIPTIONAL LEVEL

Although not intending to be limited to any theory or explanation regarding the mechanism of *SCR* action, the cloning of the gene and the expression pattern provide some clues as to the role of *SCR* in the regulation of a specific asymmetric division. The *SCR* gene is expressed in the cortex/endodermal initial, but immediately after division is restricted to the endodermal lineage. A similar pattern is seen in the ET199 enhancer trap line in which *SCR* regulatory elements are in proximity to a GUS gene, indicating that *SCR* restriction to the endodermal cell file is due to differential regulation of expression of the *SCR* gene in this cell and the first cell in the cortex file. Another marker line in which expression of GUS is detected only in the cortex daughter cell provides a control for differential degradation of GUS RNA or protein. Thus, partitioning of *SCR* RNA as a means of achieving this segregation of expression can be ruled out. What remains to be determined is whether this difference in transcriptional activity of the two daughter cells is due to internal polarity of the mother cell prior to division such that cytoplasmic determinants are unequally distributed, or to external polarity that influences cell fate after division. Since *SCR* is expressed prior to cell division, an attractive hypothesis is that it is involved in establishing polarity in the cortex/endodermal initial. The sequence of the SCR protein strongly suggests that it acts as a transcription factor. Hence, it may act to regulate the expression of other genes essential for the establishment of unequal division. Alternatively, it is conceivable that it could play a role in creating an external polarity that provides a signal to divide asymmetrically.

Its expression in more mature endodermal cells is consistent with a role in "top-down" signaling.

6.3.5. A NEW FAMILY OF TRANSCRIPTIONAL REGULATORS

Analysis of at least eighteen EST clones found in the GenBank database reveals that the proteins they encode share a high degree of homology with Arabidopsis SCR protein. See Tables 1 and 2 and FIGS. 15A-S and 28A-AH. Further sequence analysis of the encoded proteins indicate that a high degree of sequence similarity extends from at least the highly conserved VHIID domain to the carboxy-terminus of the gene products. Comparison of the amino termini of these proteins is precluded by the fact that the ESTs are incomplete. The high degree of similarity among these proteins, in combination with the motifs observed in the SCR protein (homopolymeric motifs, two leucine heptad repeats and a bZIP-like basic domain that may also function as a nuclear localization sequence) indicates that these proteins form a novel class of regulatory proteins.

The insertion sites of the T-DNA in the two *scr* mutant alleles raised the possibility that the mutant phenotype was due to the production of truncated proteins. Northern blot analysis indicated *SCR* RNA is undetectable in *scr-1*. This suggests that the phenotype is either the null, or due to highly reduced RNA expression. In *scr*-2, an alteration in RNA size was detected which would be consistent with the presence of a functional and possibly truncated protein. This could provide an explanation for the observation that *scr-2* appears to be the weaker allele.

7. EXAMPLE 2: ENHANCER TRAP ANALYSIS OF ROOT DEVELOPMENT

An enhancer trap system was used in order to provide a more detailed molecular analysis of gene expression in lateral root patterning and development in *Arabidopsis thaliana*. A new collection of marker lines that express β-glucuronidase (GUS) activity in a cell-type specific manner in each of the cells of the root was generated. These lines

allow differentiation of cells to be monitored based on molecular characteristics. One of these marker lines, ET199, resulted from the integration of the GUS cassette in proximity to a *SCR* enhancer. The results described below demonstrate that transcriptional activation of the *SCR* gene plays an important role in root development in Arabidopsis, and that *SCR* gene transcriptional regulatory elements can express a transgene in a developmentally and tissue specific manner.

7.1. MATERIALS AND METHODS

7.1.1. PLANT GROWTH CONDITIONS:

Arabidopsis seeds from NO-O and Columbia ecotypes were sterilized and sown on MS plates containing 4.5% sucrose. Plates were oriented vertically and maintained under an 18 hours light, 6 hours dark cycle.

7.1.2. HISTOLOGY AND GUS STAINING:

For observation of lateral roots, roots were removed from plates and infiltrated in 25% glycerol for several hours to overnight. Roots were then mounted in 50% glycerol. Whole seedlings were stained for GUS activity for up to three days in the following solution: 1X GUS buffer, 20% methanol, 0.5 mg/ml X-Glu. Addition of methanol greatly improves the specificity and reproducibility of staining. Staining solution was made fresh from a 10X buffer (1 M Tris pH7.5, 290 mg NaCl, 66 mg $K_3Fe(CN)_6$) that was stored for no more than one week. Stained roots were cleared in glycerol and mounted as above. All samples were observed using Nomarski optics on a Leitz Laborlux S microscope. Photographs were taken using a Leitz MPS52 camera, and images were scanned into Adobe Photoshop to create figures. In some cases the intensity of the blue color was increased.

7.1.3. CONSTRUCTION OF ENHANCER TRAP LINES:

Plant Cloning Vector (PCV) (Koncz et al., 1994, Specialized vectors for gene tagging and expression studies, in Plant Molecular Biology Manual, Gelvin & Schilperoort, eds., Vol. B2, pp. 1-2, Kluover Academic Press, Dordrecht, The Netherlands) contains a Bam HI site immediately adjacent to the T-DNA right border sequence. The β-glucuronidase gene fused to the TATA region (-46 to 78) of the CaMV 35S promoter was introduced into this site (Benfey et al., 1990, EMBO J. 9:1677-1684). 350 transgenic lines were generated by *Agrobacterium* mediated root transformation (Marton & Browse, 1991, Plant Cell Reports 10:235-239), and 4 independent lines from each transformant were screened for GUS activity in the root.

7.2. RESULTS

7.2.1. DIFFERENTIATION IN THE LRP

The marker lines described above reflect patterns of gene expression that are specific to individual root cell types. There are no readily apparent mutant phenotypes in any of these lines. Therefore, they can be used to analyze the differentiation state of the cells during normal development of the lateral root primordial (LRP). If there are stages at which the pericycle cells proliferate in the absence of patterning, it can be expected that all cells would be identical with none expressing differentiated characteristics. In contrast, organization of the LRP would be reflected in differential patterns of GUS gene expression, with certain cells beginning to turn on transcription from differentiated cell-type specific promoters (*i.e.*, those that drive GUS expression in the enhancer trap lines).

The process of lateral root formation is divided into the following seven stages:

Stage I: The LRP is first visible as a set of pericycle cells that are clearly shorter in length than their neighbors, having undergone a series of anticlinal divisions. Laskowski et al., 1995, Dev. 121:3303-3310 predict that there are approximately 4 founder pericycle cells involved. In the longitudinal plane, these divisions result in the formation of 8-10 small cells, which enlarge in a radial direction.

Stage II: A periclinal division occurs that divides the LRP into two layers (Upper Layer (UL) and Lower Layer (LL)). Not all the small pericycle-derived cells appear to participate in this division -- typically the most peripheral cells do not divide. Hence, as the UL and LL cells expand radially, the domed shape of the LRP begins to appear.

Stage III: The UL divides periclinally, generating a three layer primordium comprised of UL1, UL2 and LL. Again, some peripheral cells do not divide, creating peripheral regions that are one and two cell layers thick. This further emphasizes the domed shape of the LRP.

Stage IV: The LL divides periclinally, creating a total of four cell layers (UL1, UL2, LL1, LL2). At this stage, the LRP has penetrated the parent endodermal layer.

Stage V: The central cells in LL2 undergo a number of divisions that push the overlying layers up and distort the cells in LL1. These divisions are difficult to visualize at this stage, but clearly form a knot of mitotic activity. The LRP at this stage is midway through the parent cortex. The outer layer contains 10-12 cells.

Stage VI: This stage is characterized by several events. The four central cells of UL1 divide periclinally. This division is particularly useful in identifying the median longitudinal plane in the enlarging LRP. At this point, there are a total of twelve cells in UL1, four in the middle

that have undergone the periclinal division and four on either side. In addition, all but the most central cells of UL2 undergo a periclinal division. At this point the LRP has passed through the parent cortex layer and has penetrated the epidermis. The central cells apparently derived from LL2 have a distinct elongated shape characteristic of vascular elements.

Stage VII: As the primordium enlarges, it becomes difficult to characterize the divisions in the internal layers. However, the cells in the outermost layer can still be seen very clearly. All of these cells undergo an anticlinal division, resulting in 16 central cells (8 cells in each of two layers) flanked by 8-10 cells on each side. We refer to this as the 8-8-8 cell pattern. The LRP appears to be just about to emerge from the parent root.

7.2.2. MARKER LINES

An enhancer trapping cassette was generated by fusing the GUS coding sequence to the minimal promoter of the 35S promoter from CaMV. This minimal promoter does not produce a detectable level of GUS expression. However, its presence allows other upstream elements to direct GUS expression in a developmental and/or cell-specific manner (Benfey et al., 1990, EMBO J. 9:1677-1684). The use of a minimal promoter instead of a promoterless construct allows GUS expression to occur even if the enhancer trap cassette inserts at a distance from the coding region. Since the insert does not have to be within the structural gene, there are often no mutations generated in the enhancer trap lines. The minimal promoter:GUS construct was cloned immediately adjacent to the T-DNA right border sequence of PCV (Koncz et al., *supra*) and introduced into Arabidopsis. 350 independent lines were generated and analyzed for GUS activity in the root. The following lines most clearly define each cell type. All of the lines were generated through enhancer trapping, as described herein, below, except for CorAX92

(Dietrich et al., 1992, Plant Cell 4:1371-1382) and EpiGL2:GUS (Masucci et al., Dev. 122:1253-1260) which are transgenic plants that contain cell-type specific promoters fused to the GUS gene.

Ste05 - expresses GUS in the stele including the pericycle layer throughout primary and lateral roots. At the root tip, staining becomes weaker in the elongation zone; therefore, it is likely that only differentiated stele cells express GUS activity. Stelar GUS expression is seen also in aerial parts of the plant.

End195 - expresses GUS in the endodermis of primary and lateral roots. Staining can be seen most clearly in the cells in the meristematic region of the root, although overstaining shows that more mature cells also express some GUS activity. It appears that there is no staining in the cortex/endodermal initial, but staining is evident in the first daughter cell of this initial. GUS expression is seen also at the base of young leaves and in the stipules.

ET199 - expresses GUS in the endodermis of primary and lateral roots, again most clearly in cells in the meristematic region. Unlike End195, staining in ET199 appears to continue down to the cortex/endodermal initial and, in younger roots, even into the cells of the quiescent center. Expression in the aerial parts of the plant is detectable in the young leaf primordia.

CorAX92 - This line was generated by fusing the 5' and 3' sequences from a cortex specific gene isolated from oilseed rape to the GUS reporter gene (Dietrich et al., Plant Cell 4:1371-1382). Expression is limited to the cortex layer, extending to, but not including, the cortex/endodermal initial. Staining is also apparent in the petioles and leaf blades of expanded leaves.

EpiGL2:GUS - This line was generated by fusing the GL2 promoter to the GUS gene (Masucci et al., Dev. 122:1253-1260). Expression is seen in the non-hair forming epidermal cells (atrichoblasts). Staining is seen near the root tip, but it is difficult to determine if it includes the epidermal initial. Staining is seen also in the trichomes, leaf primordia and the epidermis of the hypocotyl and leaf petioles.

CRC219 - This line shows staining in the columella root cap only.

LRC244 - This line shows staining in the lateral root cap only.

RC162 - This line shows staining in both the lateral and columella root caps.

Two marker lines show differential staining at very early stages of LRP development. One of these, ET199, presents a complex and dynamic pattern of expression. Staining is first apparent at stage II in only the four central cells of the UL. At stage III, staining is strongest in the central cells of UL2. As the LRP reaches stage V, the staining remains strongest in the central 2-4 cells of UL2. By stage VI, staining also begins to extend into the newly formed endodermal layer, and staining in both the central cells and endodermis persists beyond emergence of the lateral root.

Another line, LRB10 (lateral root base), does not express GUS in the primary root tip. Staining in the LRP is seen at stage I, and at stage II all the cells of the UL and LL are stained. However, by stage IV and V only, the cells at the periphery of the LRP still are expressing GUS. As the LRP develops, these cells continue to stain, although less intensely, resulting in a ring of GUS expressing cells at the base of the LR.

LRB10 and ET199 clearly demonstrate non-identity between the cells at very early stages, stage IV in the case of LRB10 and within the UL at stage II in ET199. In addition, although it is difficult to identify the nature of the cells that correspond to the observed staining pattern in LRB10 and the early staining cells of ET199, post-emergent lateral roots show analogous staining in these lines, suggesting that the stained cells already are expressing markers that reflect their differentiated cell fates. Hence, these observations suggest a very early onset of differentiation in the cells of the LRP.

7.2.3. ET199 PROVIDES EVIDENCE FOR THE ROLE OF *SCR* IN PLANT DEVELOPMENT

Fortuitously, it was discovered that the GUS cassette in ET199 described Section 7.2.2, above, is situated approximately 1 kb upstream from the *SCR* gene. The *SCR* cDNA was labelled and used to probe genomic DNA from WT and ET199 plants. The band pattern seen in the Southern was completely consistent with a T-DNA inserted 1 kb upstream of the putative *SCARECROW* start site. Subsequently, a DNA fragment was PCR amplified using a primer within the T-DNA and a primer within *SCARECROW*. The size of this fragment was consistent also with the predicted insertion site. Partial sequencing of the PCR fragment confirmed the presence of *SCARECROW* sequence. Mutants in the *SCR* gene are completely lacking one of the radial layers between the epidermis and pericycle in both primary and lateral roots, due to the absence of specific cell division during embryogenesis and of the cortex/endodermal initial during post-embryonic growth. The expression pattern (described in Section 7.2.2., above) that was observed in the central cells of the developing LRP of ET199 provides strong evidence that the cells in this region are involved in the establishment of the meristematic initials. More importantly, these results demonstrate that transcriptional activation of the *SCR* gene plays a major role in the development of the Arabidopsis LRP. Furthermore,

these results demonstrate that a transgene can be expressed under the control of *SCR* gene transcriptional regulatory elements in a developmental and tissue-specific manner.

8. EXAMPLE 3: ACTIVITY OF ARABIDOPSIS *SCR* PROMOTER IN TRANSGENIC ROOTS

The expression pattern of Arabidopsis *SCR* has been determined by analysis of an enhancer trap line, ET199, in which a GUS coding region with a minimal promoter was fortuitously inserted 1 kb upstream of the *SCR* coding region (see *supra*). In ET199 plants, GUS expression is detected in the endodermis, endodermal initials and sometimes in the quiescent center (QC) of the root. See *supra* and Malamy and Benfey, 1997, Dev. 124:33-44. This expression pattern of *SCR* in the primary root has been confirmed by *in situ* analysis (See *supra* and Di Laurenzio et al., 1996, Cell 86:423-433).

The following experiments demonstrate that 2.5 kb of 5' sequence upstream of the Arabidopsis *SCR* coding region is sufficient to confer *SCR* expression pattern to a heterologous gene. The 5' sequence used in these studies starts from the Hind III site approximately 2.5 kb upstream of the ATG initiation site and extends 3' downstream to the base pair immediately upstream of the ATG initiation site (see FIG. 14). This 5' sequence was fused to a GUS coding sequence. The resulting *SCR* promoter::GUS construct was incorporated into an *Agrobacterium* vector, which was used to transform and generate transgenic roots using standard procedures.

A large number of roots were regenerated. They show GUS staining pattern that is similar to the *SCR* expression pattern in ET199 plants (Figure 19, Panel f). Since organs regenerated from callus often have an abnormal morphology, transgenic roots were transferred to liquid culture. Roots grown in liquid culture appeared morphologically normal and showed GUS expression in the endodermis, endodermal initial and QC (Figure 19, Panel g), similar to the expression pattern of *SCR* seen in the

enhancer trap line ET199. These results indicate that the 2.5 kb region upstream of the *SCR* start site is sufficient to confer the *SCR* expression pattern in the root.

The expression of the *SCR* promoter::GUS construct was examined also in the *scr* mutant background. The *scr* mutant has an altered root organization (see, *supra*). Whereas the wild-type root of Arabidopsis has four distinct cell layers surrounding the vascular tissue, the roots of *scr* mutant have only three.

Transgenic roots of the *scr* mutant that contained a *SCR* promoter::GUS construct were generated. As in the wild-type, a large number of transgenic roots were formed that had detectable GUS expression (Figure 20, Panel a). These roots were shorter than wild-type regenerated roots, consistent with the shorter root phenotype of the *scr* mutant.

Additional transgenic root experiments demonstrated that the *SCR* gene under control of its own promoter can rescue the *scr* mutant phenotype. Transgenic *scr* roots were generated that contained the full length *SCR* gene under the control of its own promoter. The length of transgenic roots containing the construct were longer than those of the *scr* mutant, indicating that the introduced *SCR* gene partially rescued the mutant. Whereas *scr* regenerated roots that carried the *SCR* promoter::GUS construct were very short (Figure 21, Panel a; and Figure 20, Panel a), roots transformed with the *SCR* promoter and coding region were noticeably longer (Figure 21, Panel b). The difference was even more obvious in liquid culture, in which *scr* mutant roots remained short (Figure 21, Panel c), while *SCR* gene complemented *scr* mutant roots were long and resembled wild-type roots (Figure 21, Panel d).

Anatomical studies of the regenerated roots confirmed the ability of the *SCR* promoter::*SCR* gene construct to rescue the *scr* mutant phenotype. Whereas regenerated roots of *scr* mutants were missing an internal layer (Figure 21, Panel e), the *scr* mutant roots that were transformed with the *SCR* promoter::*SCR* gene construct had a radial

organization that resembled wild-type root (Figure 21, Panel f).

9. EXAMPLE 4: ISOLATION OF *SCR* SEQUENCES USING PCR-CLONING STRATEGY

Based on the comparison of the sequences of *SCR* paralogs in Arabidopsis, degenerate primers SCR3AII, SCR5AII and SCR5B were designed and used in PCR amplification of *SCR* sequences from genomic DNA of various plant species. The amplification was performed according to conditions described in Section 5.1.1., *supra*, using DNA isolated from maize plants grown from a commercial seed mixture. Amplification products (104 bp fragment for the SCR5B+SCR3AII primer combination; 146 bp fragment for the SCR5AII+SCR3AII primer combination) were obtained, and each cloned into a T/A vector (Invitrogen, San Diego, CA) and sequenced. Two of the three different types of clones obtained had deduced amino acid sequences that were very similar to a part of the Arabidopsis SCR protein (*i.e.*, approximately 90% identity), suggesting that they represent parts from two different alleles of the maize *SCR* gene (*i.e.*, *ZCR* gene). The two clones each had only two conservative changes in their nucleotide sequence.

The 146 bp amplification product, ZmScl1, was subsequently used as a probe for screening of a genomic library generated in lambda BlueSTAR vector (NOVAGEN) from maize (HiII line) genomic DNA. The screening was performed according to the standard procedures described in Genius™ System User's Guide For Membrane Hybridization (Boehringer-Mannheim): The probe was a single-strand DNA molecule corresponding to the ZmScl1 fragment produced by PCR (Genius, Boehringer-Mannheim). Hybridization was performed according to recommendations of the manufacturer's manual (Boehringer-Mannheim). Prehybridization was for 2 hr in 50% formamide hybridization solution at 42°C. Hybridization was overnight at 42°C with 200 ng/ml probe concentration. Filters were washed twice at room temperature in 2x SSC, 0.1%

SDS for 5 min, and for stringent washing at 65°C in 0.5x SSC, 0.1% SDS twice for 15 min.

A positive clone was identified. The clone contained a 13 kb insert, which was subcloned into a plasmid vector. The resulting plasmid was designated pZCR. A 5 kb Eco RI fragment containing the maize *SCR* (*ZCR*) sequence was subcloned and sequenced. The nucleotide sequence of the region containing a partial *ZCR* coding sequence is shown in FIG. 17A and the corresponding deduced amino acid sequence is shown in FIG. 17B. The ZCR protein contains a segment that is highly homologous to a corresponding segment in the Arabidopsis SCR protein (FIG. 17B). This segment is flanked by segments of low homology. Thus, it is possible that the genomic clone of ZCR is a composite clone, containing sequences that are not ZCR sequences.

The deduced ZCR protein sequence was aligned with that of Arabidopsis SCR protein. The comparison revealed new conserved sites in the *SCR* coding sequence which were used to design new, more specific PCR primers (*i.e.*, 1F, 1R, and 4R) for use in amplification of *SCR* sequences from yet other plant species.

Using combinations of primers 1F+1R and 1F+4R, PCR amplification was performed as described in section 5.1.1.. Two DNAs of expected size were obtained from soybean: a 247 bp DNA from the 1F+1R primer combination and a 379 bp DNA from the 1F+4R primer combination. A DNA of expected size (247 kb) was obtained from carrot and spruce when their genomic DNA was amplified using the 1F+4R primer combination. The nucleotide sequences of the 379 kb soybean DNA (*SCLg1*), the 247 kb DNA from carrot (*SCLd1*) and spruce (*SCLp1*) are shown in FIGS. 16K-M. The corresponding deduced amino acid sequences of these amplified sequences are shown in FIG. 18. Comparison of these partial *SCR* coding sequences indicate this approach isolated DNA sequences that encode SCR proteins with amino acid sequences that are very similar, but not identical, to a segment of Arabidopsis SCR protein (see FIG. 18).

EXAMPLE 5. EXPRESSION PATTERN OF MAIZE *ZCR* GENE IN ROOT TISSUE

These experiments examined the expression pattern of *ZCR* in the primary root and quiescent centers of maize root. The expression pattern was determined by *in situ* hybridization using a *ZCR* RNA probe, corresponding to an amino acid segment region that is highly homologous to a corresponding segment of the Arabidopsis SCR protein. The experiment was carried out as follows. Restriction fragments containing the maize *ZCR* sequence were isolated from pZCR and subcloned into a pBluescript vector for *in vitro* transcription. The probe was synthesized using conditions described in the Genius Dig RNA labeling kit. The pBluescript plasmid was linearized, and 1 μg was used as a template to synthesize digoxigenin-labeled RNA using the T7 polymerase. The RNA probe was subjected to mild alkali hydrolysis by heating at 60°C for 1 hr in 100 mM carbonate buffer (pH 10.2) to yield a probe size of approximately 0.15 kb. Probe concentration for hybridization was optimized at 1 μg/ml/kb. *In situ* hybridization of root tips from 48 to 72 hr-old maize seedlings or excised quiescent centers (QCs) of roots were carried out following procedures described in Section 6.1.6., *supra*.

The results show that *ZCR* expression in maize primary roots is localized to a file of cells that is identified as the endodermal layer. The expression pattern continues in a single uninterrupted file through the QC which consists of approximately 1000-1500 cells (FIG. 22).

In two-week old regenerating QCs, *ZCR* expression is found in a file of cells extending through the newly formed apex. Thus, the regenerated roots exhibit a *ZCR* expression pattern that is similar to that seen in the primary root, even though the root apex does not contain the normal arrangement of cell files at this stage.

ZCR expression during regeneration of the root apex also was examined. In the initial stages of regeneration, cell proliferation occurs to fill in the

removed tissue and begins to regenerate the basic shape of the root tip. All cells on the blunt edge of the root appear to contribute to the new population of cells. The *ZCR* expression pattern indicates that molecular signals are differentially present in these cells at an early stage in regeneration. The gene appears to be diagnostic of cells that are preparing to undergo asymmetrical division in order to re-establish the normal organization of the root apex from the large undifferentiated cells. The results indicate that *ZCR* expression is required for pattern formation since it is expressed prior to the generation of any specific anatomical pattern in the newly formed root tissue.

11. EXAMPLE 6. EXPRESSION PATTERN OF *SCR* GENE IN SOYBEAN ROOTS AND ROOT NODULES

SCR expression in soybean roots and nodules was examined using *in situ* hybridization with a *SCR* probe. The procedures used were as described in Sections 6.1.6. and 10.

In primary roots, *SCR* is expressed in the endodermis. Expression was found also in cells at the root tip that are located at the distal end of the endodermal cell files. In soybean nodules, expression of *SCR* was detected in the peripheral tissue at the site of developing vascular strands. At later stages of vascular development within the nodule, *SCR* expression was found flanking the vascular tissue. These results indicate that *SCR* is involved in regulating vascularization in the nodule by contributing to the radial organization that is required to generate endodermis. These findings indicate that the *SCR* promoter may be used to express proteins in a highly tissue-specific manner in soybean nodules. One application is to use the *SCR* promoter to engineer nodules through production of components in a tissue-specific manner. Another application is that modification of the expression of *SCR* could enhance nodule activity by improving vascularization and/or the number of endodermal layers.

12. EXAMPLE 7. *SCR* EXPRESSION AFFECTS GRAVITROPISM OF AERIAL STRUCTURES

In addition to being defective in specific embryonic and postembryonic meristematic divisions, both the *scr* and the *shr* mutants have shoots that exhibit severely defective gravitropism. Complementation analysis showed that *scr* is allelic to a *sgr* (shoot gravitropism) mutant, *sgr1*. Four mutant alleles of *SCR* (*i.e.*, *scr1*, *scr2*, *sgr1-1* and *sgr1-2*) have been identified. All four of these mutants have normal root gravitropism and defective shoot gravitropism.

Etiolated hypocotyls of *scr* mutants placed on their sides do not respond to gravity even after 3 hr. Similar behaviors were observed with the inflorescence stems of *sgr1-1* mutant, which do not curve upwards even after two days on their sides. In contrast, the roots of these plants respond rapidly to the change in orientation with the same kinetics as the wild type. Thus, mutations in the *SCR* gene lead to a radial pattern deficiency in the root but have no effect on root gravitropism.

Comparable results were obtained also for *shr* roots and for hypocotyls and inflorescence stems, *i.e.*, data indicate that *shr* shows normal root gravitropism but almost no stem gravitropism.

13. EXAMPLE 8. MAIZE *ZCR* GENE

This example describes the cloning and expression pattern of the maize *ZCR* gene, an ortholog of the Arabidopsis *SCR* gene.

13.1. CLONING THE MAIZE *ZCR* GENE

In order to clone the maize ortholog of the Arbidopsis *SCR* gene, a reverse genetic technology strategy was utilized. With this strategy, it is possible to clone genes from across taxonomic boundaries, such as from genes identified in model organisms like Arabidopsis to those embedded in more complex genomes such as maize.

More specifically, using the deduced amino acid sequence of the Arabidopsis *SCR* gene in the reverse genetic technology strategy, multiple maize EST sequences related to *SCR* were isolated. One of them appeared very homologous to *SCR*, having greater than 77% sequence identity.

Using this highly homologous EST sequence as a probe, three genomic clones from a B73 inbred maize genomic library were isolated. Based upon restriction enzyme analysis, the three genomic clones appeared to be overlapping portions of the same genomic region.

Subsequently, a 5kb SalI fragment from one of the three clones was subcloned into pBluescript SK(-) and sequenced. The sequence analysis of the cloned maize gene revealed that it consists of two exons and one intron in one open-reading frame (ORF) encoding 668 amino acids. The presence of an in-frame stop codon located 5' to the initiating ATG and nearby stop codons in the other two reading frames suggests that the long ORF of this genomic clone encodes the functional, full length protein. *See*, FIG. 25.

After obtaining the full length maize sequence, a database search was performed to find homologous sequences. The database search revealed that the newly isolated maize sequence was most homologous with the Arabidopsis *SCR* gene at the amino acid level. Comparison of the maize and Arabidopsis sequences indicated that the similarity between the Arabidopsis *SCR* and the maize *ZCR* gene extended beyond the VHIID domain into both the N- and C-termini (FIG. 26). Although the N-terminal region of the maize ortholog and the Arabidopsis *SCR* gene appears more divergent, the maize *ZCR* gene has the homopolymeric stretches characteristic of *SCR* (Gerber et al., 1994, Science 263:808-811; Johnson, et al., 1993, J. Nutr. Biochem. 4:386-398).

In addition, the *ZCR* gene has other motifs characteristic of *SCR*: two putative leucine heptad repeats, which have been shown in other proteins to mediate protein-protein interactions; and a stretch of basic residues

similar to the basic domain of bZIP proteins, which have been shown not only to mediate DNA-binding, but also nuclear localization (Hurst, H.C., 1994, Protein Prof. 1:123-168). Moreover, the *ZCR* gene has three copies of an LXXLL motif in the N-terminal region, which has been shown to mediate the binding of a steroid receptor coactivator complex to nuclear receptors (Heery, et al., 1997, Nature 387:733-736; Torchia, et al., 1997, Nature 387:677-684). *See*, FIG. 26. Similarly, the GAI and RGA gene products also contain a copy of this sequence. In these genes, the sequence is believed to be involved in a gibberellin signal transduction pathway (Peng, et al., 1997, Genes Dev. 11:3194-3205; Silverstone, et al., 1998, Plant Cell 10:155-169).

Although the functionality of these putative motifs has not been clearly demonstrated, the fact that all of these putative motifs exist in a single polypeptide strongly suggests that the maize *ZCR* is a transcription factor similar to the Arabidopsis *SCR* gene. In addition, the structure of the *ZCR* gene is very homologous to that of the *SCR* gene. Specifically, the position of the intron is conserved, although the size and sequence of the intron is different in the two genes.

In addition to the maize *ZCR* gene, a 3.2kb fragment upstream of the initiating ATG of the maize gene was isolated. This region, similar to numerous other upstream regions in other genes, likely contains regulatory elements of the *ZCR* gene. Furthermore, this upstream region can be analyzed and utilized similar to the upstream region of the *SCR* gene, discussed *supra*.

FIG. 32 shows an RNA blot analysis in which either total RNA or poly-A selected RNA from roots and shoots were probed with the full-length *ZCR* cDNA. As shown in the figure, the probe hybridized to a band that is approximately 2.6 kilobases in size.

FIG. 33 shows the partial nucleic acid and amino acid sequence of CBPBT44, a gene which has significant homology to both the Arabidopsis *SCR* and the maize *ZCR* genes.

FIG. 34 represents an alignment of the three genes. As shown in FIG. 34, the three genes share a high degree of homology, including, but not limited to, the leucine heptad repeats. To further demonstrate the homology between the maize *ZCR* gene and the CBPBT44 partial sequence, a Southern blot analysis was performed. *See*, FIG. 35. FIG. 35 demonstrates that CBPBT44 (right pane, lane C) is the source of some of the bands picked up by the maize *ZCR* cDNA (right panel, lane A). Thus, it is likely that CBPBT44 is a closely related gene to the *ZCR* gene, and that CBPBT44 may represent a duplicated copy of the maize *ZCR* gene in the maize genome. This possibility is strengthened by the fact that maize is thought to have undergone a general duplication of its genome during its evolution.

13.2. EXPRESSION PATTERN OF THE MAIZE *ZCR* GENE

In order to understand the function of the maize *ZCR* ortholog, the expression pattern of the maize ortholog was examined in various types of roots, including, but not limited to, the maize primary, embryonic, lateral, seminal lateral and adventitious roots by RNA *in situ* hybridization. Suprisingly, in spite of the profound differences of the root architecture between maize and Arabidopsis (FIG. 23), the expression pattern of the maize *ZCR* is remarkably similar to that of the Arabidopsis *SCR* in that expression is found only in the endodermis cell lineage (Fig. 22A-C). Furthermore, it is expressed in the embryonic root and lateral root (FIG. 22D-F).

Interestingly, *ZCR* expression also was found to extend through the QC (FIG. 22A-B). Expression through the QC was confirmed by observations of the expression pattern in serially cut sections. This demonstrates the first evidence for cell-specific expression within the QC, which has long been considered to be undifferentiated and probably

multipotent, analogous to stem cells in animals (Barlow, P.W., 1976, J. Theor. Biol. 57:433-451; Barlow, P.W., 1978, In Stem cells and tissue homeostasis (Lord, B. I., Potten, C. S. and Cole, R. J. eds), (Cambridge: Cambridge University Pess)). In addition, this finding raises the possibility that radial organization is established in the mitotically inactive narrow region where cell files converge.

14. EXAMPLE 9: MAIZE *ZCR* GENE EXPRESSION DURING REGENERATION OF THE ROOT TIP

This example describes the expression of the maize *ZCR* gene during regeneration of the root apex after excision of the QC. Expression after removal of the root cap and immediately after QC excision did not show any alteration in its pattern (FIGS. 27A-B).

At 24 hours after removal of the QC, the excised tissue began to be replaced, reforming the basic shape of the root tip. Expression was found in the endodermal cell file of the unexcised portion of the root as well as in the newly formed cells at the base of the endodermal cell files. The lack of its expression in the cells below this region indicates that it is activated only after initial proliferation and partial restoration of the apex. Moreover, expression was found also in isolated cells located between the cell files (FIG. 27C). Examination of serially cut transverse sections indicated that these internal cells were not directly adjacent to any other cells expressing the gene (FIG. 27D). This observation indicates that there is no lineage requirement for the isolated cells expressing the maize *ZCR* gene.

At 48 hours after excision of the QC, expression of the maize *ZCR* was found in a band of cells that is nearly perpendicular to the base of the endodermal cell files (FIG. 27E). At this stage, the root tip had regained its normal external shape, although longitudinal sections show that the cell files are not organized into the converging files seen in the normal root anatomy.

At 72 hours, the expression of the maize *ZCR* gene pattern resembled that found in the unexcised root, although the anatomical pattern was not yet restored (FIG. 27F). Between 72 and 96 hours, there was an anatomical shift such that files became convergent at the tip. Finally, by 96 hours following excision of the QC, *ZCR* gene expression was found to be localized to a single file of cells extending through the tip in a manner similar to that seen in the primary root (FIG. 27G).

These results show that the expression pattern of the maize ortholog converges at the root tip prior to the anatomical pattern of the root. Thus, *ZCR* gene expression prepatterns radial organization of the root. The progressive refinement of the expression pattern suggests that radial patterning is regenerated by processes that involve positional information possibly transmitted through cell-cell signaling within the regenerating region.

15. DEPOSIT OF MICROORGANISMS

The following microorganisms have been deposited in accordance with the terms of the Budapest Treaty with the American Type Culture Collection; 12301 Parklawn Drive, Rockville, MD 20852, U.S.A., on the dates indicated:

Microorganism	Clone	Accession No.	Date
DH5α	pGEX-2TK* (pLIG 1-3/Sac+MOB1Sac)	98031	April 26, 1996
DH5α	pNYH1 (Zm-scl1b)	98032	April 26, 1996
DH5α	pNYH2 (Zm-scl1)	98033	April 26, 1996
DH5α	pNYH3 (Zm-scl2)	98034	April 26, 1996
DH5α	pZCR	97992	April 18, 1997

Although the invention is described in detail with reference to specific embodiments thereof, it will be understood that variations which are functionally equivalent are within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and

described herein will become apparent to those skilled in the art from the foregoing description and accompanying drawings. Such modifications are intended to fall within the scope of the appended claims.

Various publications are cited herein, each of the disclosures of which is incorporated by reference in its entirety.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule comprising a nucleotide sequence that encodes a SCARECROW protein containing an amino acid sequence substantially similar to the sequence of MOTIF III (VHIID) of Arabidopsis SCR protein shown in FIGS. 13A-F.

2. An isolated nucleic acid molecule comprising (a) a nucleotide sequence that encodes a scarecrow protein having the amino acid sequence of SEQ ID NO:2, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:61, SEQ ID NO:63, SEQ ID NO:65, SEQ ID NO:67 or the amino acid sequence shown in FIG. 25, FIG. 28AB, FIG. 28AC, FIG. 28AD, FIG. 28AE, FIG. 28AF, FIG. 28AG or FIG. 28AH; or (b) the complement of the nucleotide sequence of (a).

3. An isolated nucleic acid molecule comprising a nucleotide sequence that hybridizes to the nucleic acid of Claim 2 and encodes a naturally occurring *SCR* gene product.

4. A nucleic acid molecule comprising (a) a nucleotide sequence that encodes a *SCR* protein lacking one to four of the following motifs delineated in FIGS. 13A-F: MOTIF I, MOTIF II, MOTIF III, MOTIF IV, MOTIF V, or MOTIF VI; or (b) the complement of the nucleotide sequence of (a).

5. A nucleic acid molecule comprising (a) a nucleotide sequence that encodes a polypeptide corresponding to MOTIF I, MOTIF II, MOTIF IV, MOTIF V or MOTIF VI of the SCARECROW protein delineated in FIGS. 13A-F; or (b) the complement of the nucleotide sequence of (a).

6. The isolated nucleic acid molecule of Claim 1 comprising the nucleic acid sequence of SEQ ID NO:1, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66 or the nucleic acid sequence shown in FIG. 25, FIG. 28A, FIG. 28B, FIG. 28C, FIG. 28D, FIG. 28E, FIG. 28F, FIG. 28G, FIG. 28H, FIG. 28I, FIG. 28J, FIG. 28K, FIG. 28L, FIG. 28M, FIG. 28N, FIG. 28O, FIG. 28P, FIG. 28Q, FIG. 28R, FIG. 28S, FIG. 28T, FIG. 28U, FIG. 28V, FIG. 28W, FIG. 28X, FIG. 28Y, FIG. 28Z or FIG. 28AA.

7. A DNA vector containing the nucleic acid molecule of Claim 1, 2, 3, 4, 5, or 6.

8. An expression vector containing the nucleic acid molecule of Claim 1, 2, 3, 4, 5, or 6 operatively associated with a regulatory sequence containing transcriptional and translational regulatory elements that control expression of the nucleotide sequence in a host cell.

9. A genetically-engineered host cell containing the nucleic acid molecule of Claim 1, 2, 3, 4, 5, or 6.

10. A genetically-engineered host cell containing the nucleic acid molecule of Claim 1, 2, 3, 4, 5, or 6 operatively associated with a regulatory sequence containing transcriptional and translational regulatory elements that control expression of the nucleotide sequence in a host cell.

11. An isolated SCARECROW protein.

12. The protein of Claim 11 having the amino acid sequence shown in FIG. 25.

13. A SCARECROW protein lacking one to four of the following motifs delineated in FIGS. 13A-F: MOTIF I, MOTIF II, MOTIF III, MOTIF VI, MOTIF V, or MOTIF VI.

14. A polypeptide corresponding to MOTIF I, MOTIF II, MOTIF IV, MOTIF V or MOTIF VI of the SCARECROW protein as delineated in FIGS. 13A-F.

15. An antibody that immunospecifically binds the protein or polypeptide of Claim 11, 12, 13 or 14.

16. An anti-idiotypic antibody that mimics an epitope of SCARECROW protein.

17. A plant genetically-engineered to overexpress or underexpress a SCARECROW protein or polypeptide, so that cell division is modified, and root and/or stem development is altered.

18. A plant genetically-engineered to overexpress a SCARECROW protein or polypeptide, so that cell division is increased in roots, resulting in thicker root development.

19. A transgenic plant containing a transgene having the nucleic acid molecule of Claim 1, 2, 3, 4, 5, or 6.

20. A transgenic plant containing a transgene having the nucleic acid molecule of Claim 1, 2, 3, 4, 5, or 6 operatively associated with a regulatory sequence containing transcriptional and translational regulatory elements that control expression of the nucleotide sequence in a transgenic plant cell.

21. The transgenic plant of Claim 19, in which the transgene encodes an antisense nucleotide sequence that suppresses expression of endogenous *SCARECROW* gene product, so that cell

division is decreased in roots, resulting in thinner root development.

22. A genetically-engineered plant in which the endogenous *SCARECROW* gene is disrupted or inactivated so that cell division is decreased in roots, resulting in thinner root development.

23. A transgenic plant containing a transgene encoding a gene of interest operatively associated with a *SCARECROW* promoter, so that the gene of interest is expressed in a tissue-specific manner in roots or embryos.

24. The transgenic plant of Claim 23, in which the gene of interest encodes a gene product that confers herbicide, salt, pathogen, or insect resistance.

25. A transgenic plant containing a transgene encoding a gene of interest operatively associated with a *SCARECROW* promoter, so that the gene of interest is expressed in shoots.

26. The transgenic plant of Claim 25, in which the gene of interest encodes a gene product that increases starch, lignin or cellulose biosynthesis.

27. A plant genetically-engineered to overexpress or underexpress the SCARECROW protein so that gravitropism of the stem or hypocotyl is altered.

28. The plant of Claim 27, which is less susceptible to lodging than a wild-type plant.

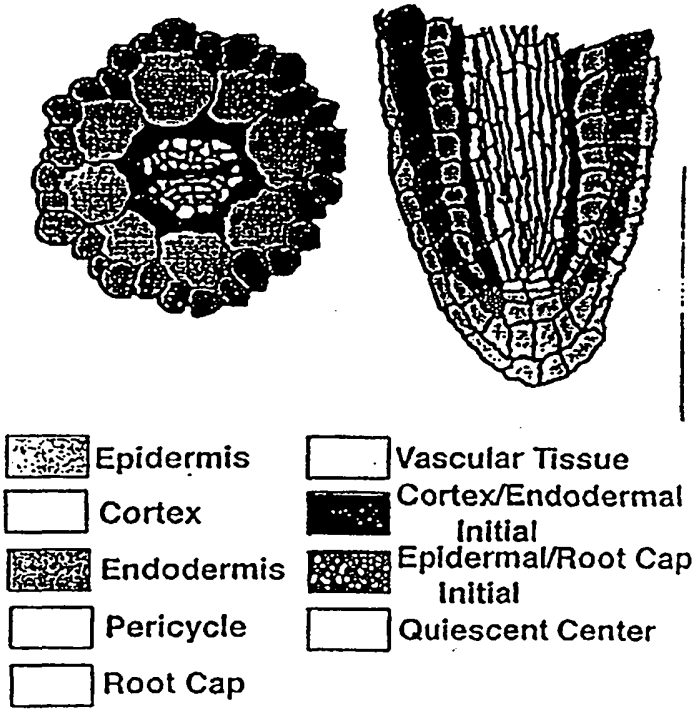

Epidermis
Cortex
Endodermis
Pericycle
Root Cap
Vascular Tissue
Cortex/Endodermal Initial
Epidermal/Root Cap Initial
Quiescent Center

FIG. 1A

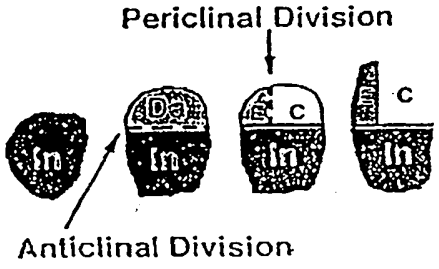

Periclinal Division
Anticlinal Division
D
C
C
In
In
In
In

FIG. 1B

FIG. 2A

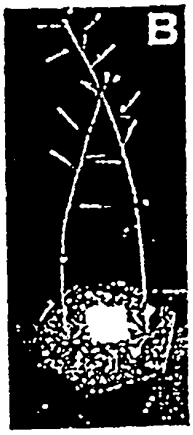


FIG. 2B

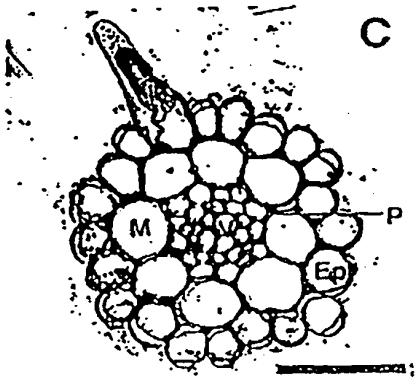


FIG. 2C

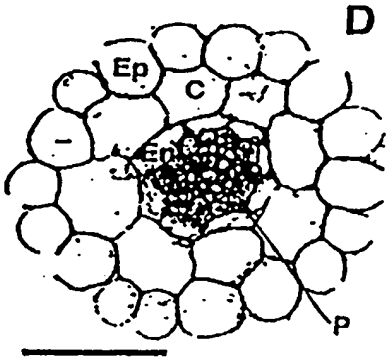


FIG. 2D

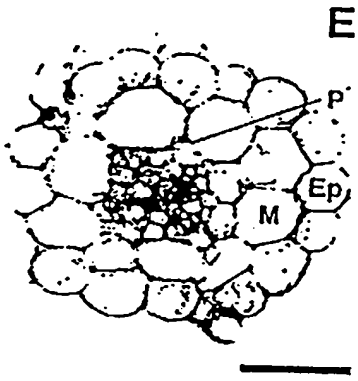


FIG. 2E

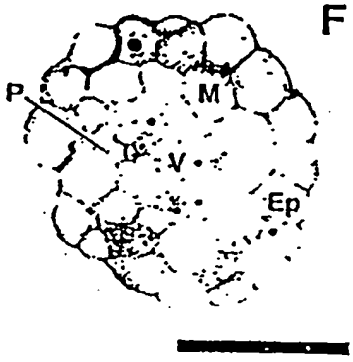


FIG. 2F

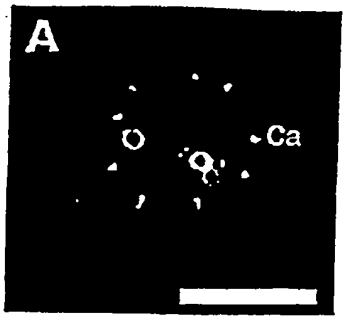


FIG. 3A

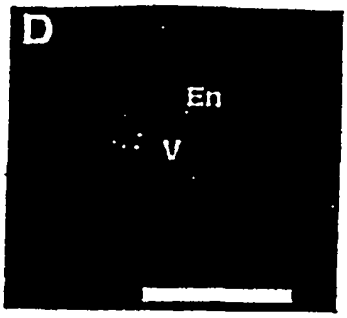


FIG. 3D

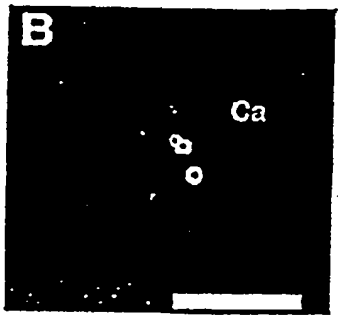


FIG. 3B

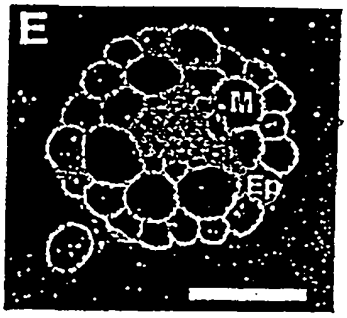


FIG. 3E

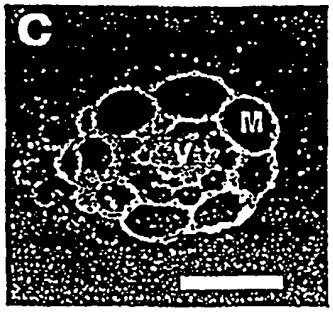


FIG. 3C

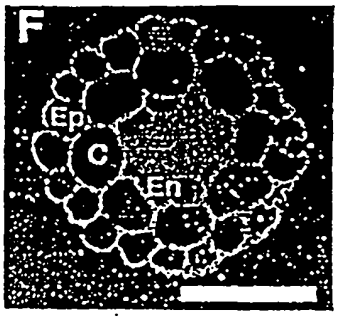


FIG. 3F

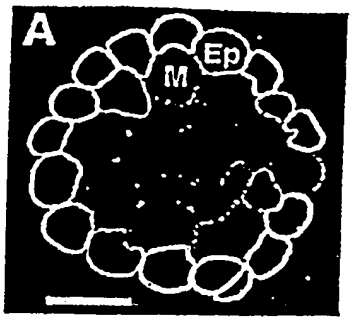


FIG. 4A

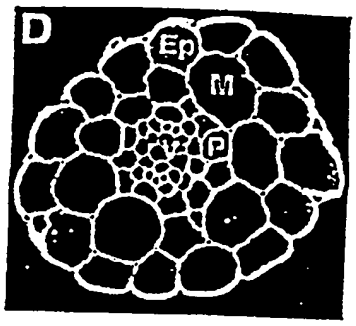


FIG. 4D

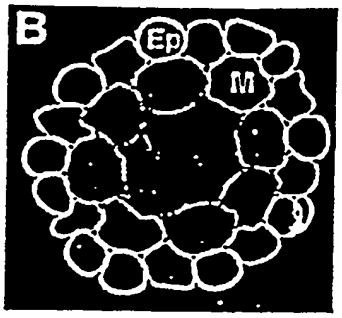


FIG. 4B

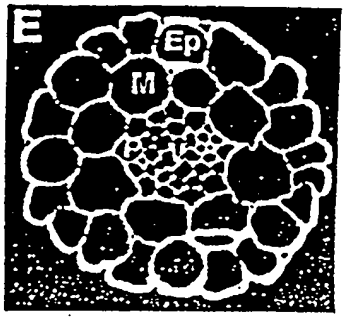


FIG. 4E

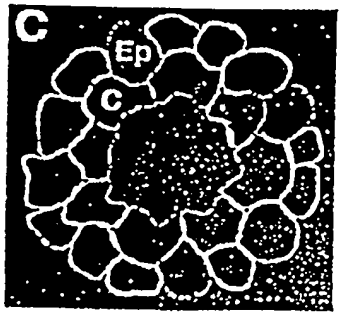


FIG. 4C

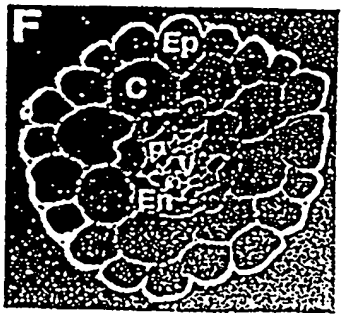


FIG. 4F

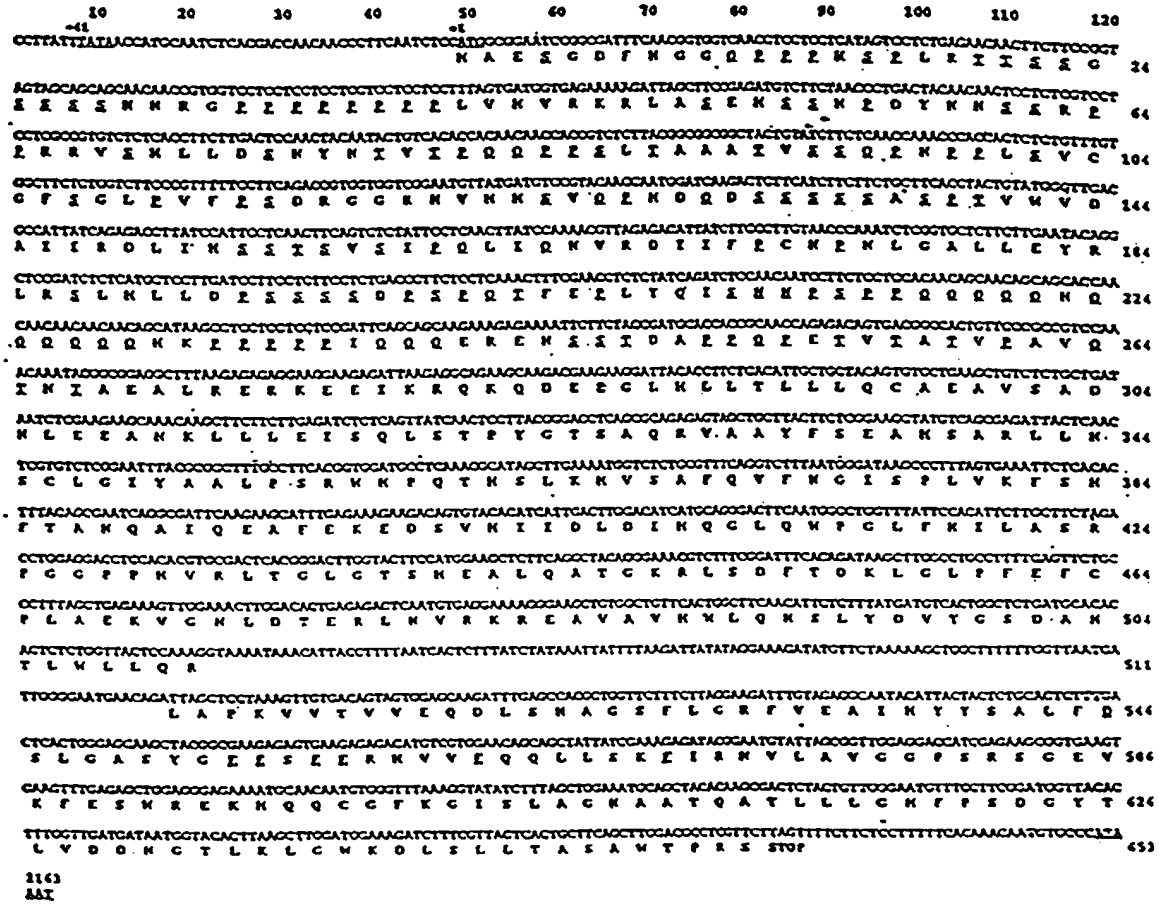


FIG. 5A

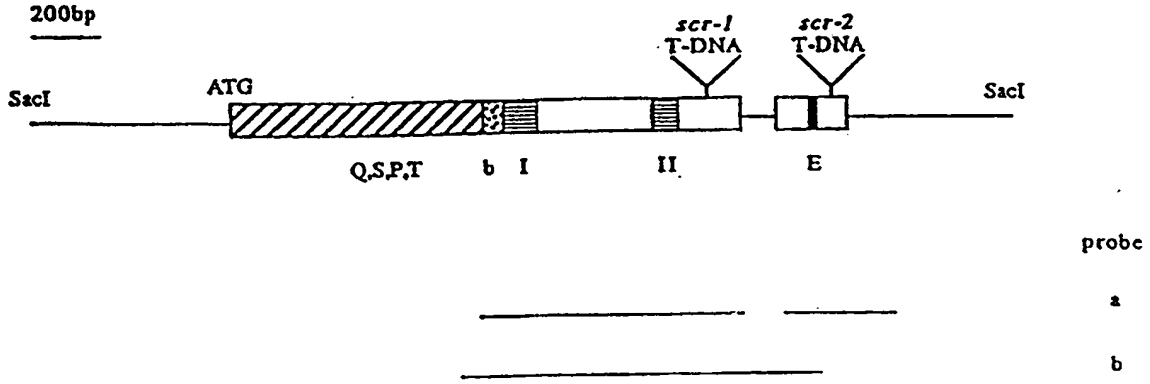


FIG. 5B

			1
SCR bZIP-like domain		PAVQTNTAEALRERKEEIKRQKQ	D
		|| | | | : | ||	
GCN4	(yeast)	LKRARNTEAARRSRARKLQRMKQ	L
TGA1	(Arabidopsis)	RRLAQNREAARKSRLRKKAYVQQ	L
C-Fos	(mouse)	IRRERNKMAAAKCRNRRRELTDT	L
c-JUN	(human)	RKRMRNRIAASKCRKRKLERIAR	L
CREB	(human)	VRLMKNREAARECRRKKKEYVKC	L
Opaque-2	(maize)	KRKESNRESARRSRYRKAAHLKE	L
OBF2	(maize)	MRQIRNRDSAMKSRERKKSYIKD	L
RAF-1	(rice)	RRMVSNRESARRSRKKKQAHLAD	L

FIG. 5C

```
SCR.VHIID domain
                                                                  1
SCR         AFEKEDSVHIIDLDIMQGLQWPGLFHILASRPGGPPHVRLTGL
F13896      AVKNESFVHIIDFQISQGGQWVSLIRALGARPGGPPNVRITGI
Z37192      AMEGEKMVHVIDLDASEPAQWLALLQAFNSRPEGPPHLRITGV
Z25645      AIKGEEEVHIIDFDINQGNQYMTLIRSIA
D41474              IHVIDFXLGVGGQWASFLQELAHRRG
T18310              VHIIXFXLMQGLQWPALMDVFSARKGGPPKLRITGI
```

FIG. 5D

MetAlaGluSerGlyAspPheAsnGlyGlyGlnProProProHisSerProLeuArgThr
ThrSerSerGlySerSerSerSerAsnAsnArgGlyProProProProProProProPro
LeuValMetValArgLysArgLeuAlaSerGluMetSerSerAsnProAspTyrAsnAsn
SerSerArgProProArgArgValSerHisLeuLeuAspSerAsnTyrAsnThrValThr
ProGlnGlnProProSerLeuThrAlaAlaAlaThrValSerSerGlnProAsnProPro
LeuSerValCysGlyPheSerGlyLeuProValPheProSerAspArgGlyGlyArgAsn
ValMetMetSerValGlnProMetAspGlnAspSerSerSerSerSerAlaSerProThr
ValTrpValAspAlaIleIleArgAspLeuIleHisSerSerThrSerValSerIlePro
GlnLeuIleGlnAsnValArgAspIleIlePheProCysAsnProAsnLeuGlyAlaLeu
LeuGluTyrArgLeuArgSerLeuMetLeuLeuAspProSerSerSerSerAspProSer
ProGlnThrPheGluProLeuTyrGlnIleSerAsnAsnProSerProProGlnGlnGln
GlnGlnHisGlnGlnGlnGlnGlnHisLysProProProProProIleGlnGlnGln
GluArgGluAsnSerSerThrAspAlaProProGlnProGluThrValThrAlaThrVal
ProAlaValGlnThrAsnThrAlaGluAlaLeuArgGluArgLysGluGluIleLysArg
GlnLysGlnAspGluGluGlyLeuHisLeuLeuThrLeuLeuLeuGlnCysAlaGluAla
ValSerAlaAspAsnLeuGluGluAlaAsnLysLeuLeuLeuGluIleSerGlnLeuSer
ThrProTyrGlyThrSerAlaGlnArgValAlaAlaTyrPheSerGluAlaMetSerAla
ArgLeuLeuAsnSerCysLeuGlyIleTyrAlaAlaLeuProSerArgTrpMetProGln
ThrHisSerLeuLysMetValSerAlaPheGlnValPheAsnGlyIleSerProLeuVal
LysPheSerHisPheThrAlaAsnGlnAlaIleGlnGluAlaPheGluLysGluAspSer
ValHisIleIleAspLeuAspIleMetGlnGlyLeuGlnTrpProGlyLeuPheHisIle
LeuAlaSerArgProGlyGlyProProHisValArgLeuThrGlyLeuGlyThrSerMet
GluAlaLeuGlnAlaThrGlyLysArgLeuSerAspPheThrAspLysLeuGlyLeuPro
PheGluPheCysProLeuAlaGluLysValGlyAsnLeuAspThrGluArgLeuAsnVal
ArgLysArgGluAlaValAlaValHisTrpLeuGlnHisSerLeuTyrAspValThrGly

SerAspAlaHisThrLeuTrpLeuLeuGlnArgLeuAlaProLysValValThrValVal
GluGlnAspLeuSerHisAlaGlySerPheLeuGlyArgPheValGluAlaIleHisTyr
TyrSerAlaLeuPheAspSerLeuGlyAlaSerTyrGlyGluGluSerGluGluArgHis
ValValGluGlnGlnLeuLeuSerLysGluIleArgAsnValLeuAlaValGlyGlyPro
SerArgSerGlyGluValLysPheGluSerTrpArgGluLysMetGlnGlnCysGlyPhe
LysGlyIleSerLeuAlaGlyAsnAlaAlaThrGlnAlaThrLeuLeuLeuGlyMetPhe
ProSerAspGlyTyrThrLeuValAspAspAsnGlyThrLeuLysLeuGlyTrpLysAsp
LeuSerLeuLeuThrAlaSerAlaTrpThrProArgSerSTOP

FIG. 5E

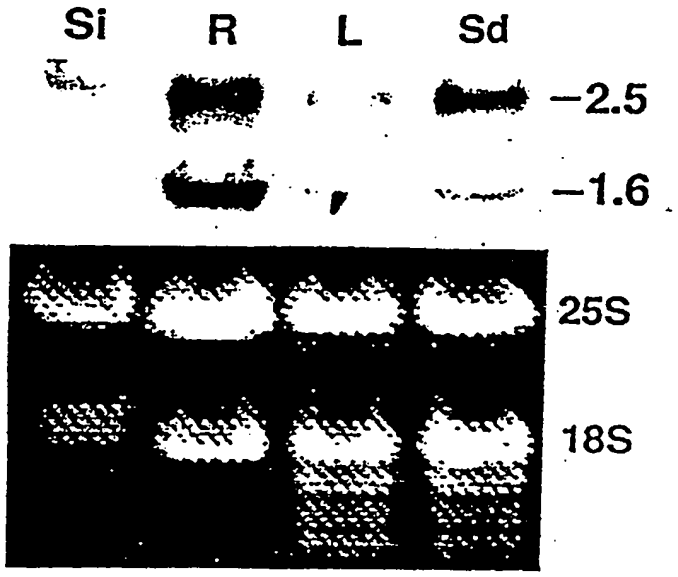


FIG. 6A

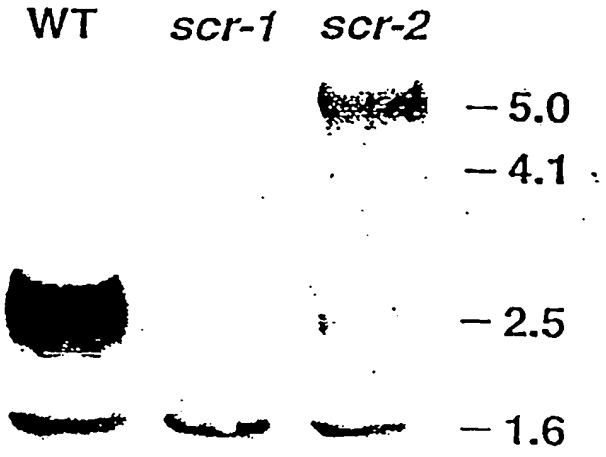


FIG. 6B

FIG. 7A

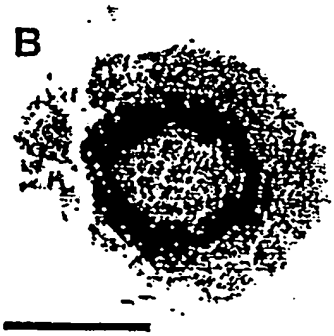


FIG. 7B

FIG. 7C

FIG. 7D

FIG. 7E

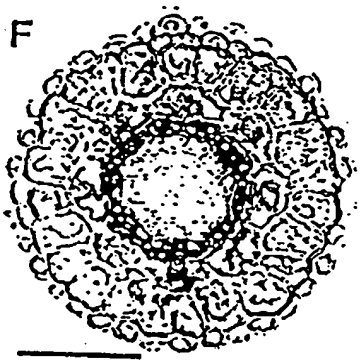


FIG. 7F

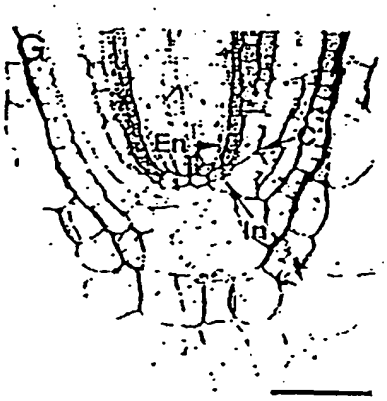


FIG. 7G

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
GGCACGAGCC CAACGGGTCC TGAGCTTCTT ACTTATATGC ATATCTTGTA     50
G T S P  T G P  E L L  T Y M H  I L Y

TGAAGCCTGC CCTTATTTCA AATTCGGTTA TGAATCTGCT AATGGAGCTA    100
 E A C  P Y F K  F G Y  E S A  N G A I

TAGCTGAAGC TGTGAAGAAC GAAAGTTTTG TGCACATTAT CGATTTCCAG    150
  A E A  V K N  E S F  V H I  I D F Q

ATTTCTCAAG GTGGTCAATG GGTGAGTTTG ATCCGTGCTC TTGGTGCTAG    200
I S Q G  G Q W  V S L  I R A  L G A R

ACCTGGTGGA CCTCCGAACG TTAGGATAAC GGGAATTGAT GATCCGAGAT    250
 P G G  P P N V  R I T  G I D  D P R S

CATCGTTTGC TCGTCAAGGA GGACTTGAGT TAGTTGGACA AAGACTTGGG    300
  S F A  R Q G  G L E  L V G  Q R L G

AAGCTAGCTG AAATGTGCGG TGTTCCGTTT GAGTTCCATG GAGCTGCTTT    350
K L A E  M C G  V P F  E F H  G A A L

ATGCTGCACG GAAGTCGAAA TCGAGAAGCT AGGAGTTAGA AATGGAGAAG    400
 C C T  E V E  I E K  L G V  R N G  E A

CGCTCGCGGT TAACTTCCCG CTTGTTCTTC ACCACATGCC TGATGAGAGT    450
  L A V  N F P  L V L  H H M  P D E  S

GTAACTGTGG AGAATCACAG AGATAGATTG TTGAGATTGG TCAAACACTT    500
V T V E  N H R  D R L  L R L  V K H  L

GTCACCAAAC GTTGTGACTC TGGTTGAGCA AGAAGCGAAT ACAAACACTG    550
 S P N  V V T  L V E  Q E A  N T N  T A

CGCCGTTTCT TCCCCGGTTT GTCGAGACAA TGAACCATTA CTTGGCAGTT    600
  P F L  P R F  V E T  M N H  Y L A  V
```

FIG. 8

```
      10        20        30        40        50
1234567890 1234567890 1234567890 1234567890 1234567890
TTCGAATCAA TAGATGTGAA ACTGGCTAGA GATCACAAGG AAAGGATCAA   650
F  E  S  I   D  V  K   L  A  R   D  H  K  E   R  I  N

TGTTGAGCAG CATTGTTTGG CTAGAGAGGT TGTGAATCTT ATAGCTTGTG   700
 V  E  Q   H  C  L  A   R  E  V   V  N  L   I  A  C  E

AAGGTGTTGA AAGAGAAGAG AGGCACGAGC CACTAGGGAA ATGGAGGTCT   750
  G  V  E   R  E  E   R  H  E  P   L  G  K   W  R  S

CGGTTTCACA TGGCGGGATT TAAACCGTAT CCTTTGAGCT CGTATGTGAA   800
R  F  H  M   A  G  F   K  P  Y   P  L  S  S   Y  V  N

CGCAACAATC AAAGGATTGC TTGAGAGTTA TTCAGAGAAG TATACACTTG   850
 A  T  I   K  G  L  L   E  S  Y   S  E  K   Y  T  L  E

AAGAAAGAGA TGGAGCATTG TATTTAGGAT GGAAGAATCA ACCTCTTATC   900
  E  R  D   G  A  L   Y  L  G  W   K  N  Q   P  L  I

ACTTCTTGTG CTTGGAGGTA ACTAATAAAA ACCTTGTTCG GTTTCAGAAG   950
T  S  C  A   W  R  X

AGATTAGAAA CTTCTTTTAA AGTTTGCAGA ATCTGTTTGT AAAAGTAAAA  1000

CTCATGCATG ATCCGNAGGA ACAAGTTGTC AAATGTTGTA GTAGTAAGTG  1050

ATATGTTGAT GACCCAAAAA AAAAAAAAAA AAAAA                  1085
```

FIG. 8 (cont'd.)

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
GCTATGGAAG GAGAGAAGAT GGTTCATGTG ATTGATCTCG ATGCTTCTGA      50
A M E G  E K M  V H V  I D L D  A S E

GCCAGCCTCAA TGGCTTGCTT TGCTTCAAGC TTTTAACTCT AGGCCTGAAG    100
 P A Q  W L A L  L Q A  F N S  R P E G

GTCCACCTCA TTTGAGAATC ACTGGTGTTC ATCACCAGAA GGAAGTGCTT    150
  P P H  L R I  T G V H  H Q K  E V L

GAACAAATGG CTCATAGACT CATTGAGGAA GCAGAGAAAC TCGATATCCC    200
E Q M A  H R L  I E E  A E K L  D I P

GTTTCAGTTT AATCCCGTTG TGAGTAGGTT AGACTGTTTA AATGTAGAAC    250
 F Q F N  P V V  S R L  D C L  N V E Q

AGTTGCGGGT TAAAACAGGA GAGGCCTTAG CCGTTAGCTC GGTTCTTCAA    300
  L R V  K T G  E A L A  V S S  V L Q

TTGCATACCT TCTTGGCCTC TGATGATGAT CTCATGAGAA AGAACTGCGC    350
L H T F  L A S  D D D  L M R K  N C A

TTTACGGTTT CAGAACAACC CTAGTGGAGT TGACTTGCAG AGAGTTCTAA    400
 L R F  Q N N P  S G V  D L Q  R V L M

TGATGAGCCA TGGCTCTGCA GCTGAGGCAC GTGAGAATGA TATGAGTAAC    450
  M S H  G S A  A E A R  E N D  M S N

AACAATGGGT ATAGCCCTAG CGGTGACTCG GCCTCATCTT TGCCTTTACC    500
N N G Y  S P S  G D S  A S S L  P L P

AAGTTCAGGA AGGACTGATA GCTTCCTCAA TGCTATTTGG GGTTTGTCTC    550
 S S G  R T D S  F L N  A I W  G L S P

CAAAGGTCAT GGTGGTCACT GAGCAAGACT CAGACCACAA CGGCTCCACA    600
  K V M  V V T  E Q D S  D H N  G S T
```

FIG. 9

```
          10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
CTAATGGAGA GGCTATTAGA ATCACTTTAC ACCTACGCAG CATTGTTTGA   650
L  M  E  R  L  L  E  S  L  Y  T  Y  A  A  L  F  D

TTGCTTGGAA ACAAAAGTTC CAAGAACGTC TCAAGATAGG ATCAAAGTGG   700
 C  L  E  T  K  V  P  R  T  S  Q  D  R  I  K  V  E

AGAAGATGCT CTTCGGGGAG GAGATCAAGA ACATCATATC CTGCGAGGGA   750
  K  M  L  F  G  E  E  I  K  N  I  I  S  C  E  G

TTTGAGAGAA GAGAAAGACA CGAGAAGCTT GAGAAATGGA GCCAGAGGAT   800
F  E  R  R  E  R  H  E  K  L  E  K  W  S  Q  R  I

CGATTTGGCT GGTTTTGGGA ATGTTCCTCT TAGCTATTAT GCGATGTTGC   850
 D  L  A  G  F  G  N  V  P  L  S  Y  Y  A  M  L  Q

AGGCTAGGAG ATTGCTTCAA GGGTGCGGTT TGATGGGTA TAGAATCAAG   900
  A  R  R  L  L  Q  G  C  G  F  D  G  Y  R  I  K

GAAGAGAGCG GGTGCGCAGT AATTTGCTGG CAAGATCGAC CTCTATACTC   950
E  E  S  G  C  A  V  I  C  W  Q  D  R  P  L  Y  S

GGTATCAGCT TGGAGATGCA GCAAGTGAAT GATATATTAC AGTTTGTCTT  1000
 V  S  A  W  R  C  R  K  X

CTATTTTGGT TATGAGCAGA GTCCCTTTCT TTTTTGTATA CATGGGGACA  1050

CAATCTTAGT TGTTTTGTGA TGGTGACTTT CTGTCTCTTT ATGCTATTTT  1100

GGCTTAAATG CTTCTACTGC CTCTGCATGT AAAGCCTTTG TGTGTTGGTT  1150

CAATTTGGTC TGGTGTGGGT GTAATACCAA ACCAAATCCA ATTTGAGCTG  1200

AAGATAACTA ATTTGATGAT CGGCTCGTGC C                      1231
```

FIG. 9 (cont'd.)

```
CTTTGTCAAT GGTAAATGAG CTGAGGCAGA TAGTTTCTAT CCAAGGAGAC      50
CCTTCTCAGA GAATCGCAGC TTACATGGTG GAAGGTCTAG CTGCAAGAAT     100
GGCCGCTTCA GGAAAATTCA TCTACAGAGC ATTGAAATGC AAAGAGCCTC     150
CTTCGGATGA GAGGCTTGCA GCTATGCAAG TCCTGTTTGA AGTCTGCCCT     200
TGTTTCAAGT TCGGGTTTTT AGCAGCTAAT GGTGCGATAC TTGAAGCAAT     250
CAAAGGTGAA GAAGAAGTTC ACATAATCGA TTTCGATATA AACCAAGGGA     300
ACCAATACAT GACACTGATA CGAAGCATTG CTGAGTTGCC TGGTAAACGA     350
CCTCGCCTGA GGTTAACAGG AATTGATGAC CCTGAATCAG TCCAACGCTC     400
CATTGGAGGG CTAAGAATCA TCAATCTAAG ACTCGAGCAA CTCGCAGAGG     450
ATAATGGAGT ATCCTTCAAA TTCAAAGCAA TGCCTTCAAA GACTTCGATT     500
GTCTCTCCAT CAACACTCGG TTGCAAACCA GGAGAAACCT TAATCAGTGA     550
ACTTTGCATT CCAACTTCAC CACATGCCTG ACGAGAGTGT CACAACAGTA     600
AACCAGCGGG ACGAGCTACT TCACATGGTC AAAAGCTTAA ACCCGCTTGT     650
CACGGTCGTT GAACAAGACG TGAACACAAA CACTTCACCG TTCTTTCCCA     700
GATTCATAGA GGCTTACGAA TACTACTCAG CAGTTTTCGA.GTCTCTAGAC     750
ATGACACTTC CAAGAGAAAG CCAAGAGAGG ATGAATGTAG AAAGACAGTG     800
TCTCGCTAGA GACATAGTCA ACATTGTTGC TTGCGAAGGA GAAGAACGGA     850
TAGAGAGATA CGAGGCTGCG GGAAAATGGA GAGCAAGGAT GATGATGGCT     900
GGATTCAATC CAAAACCAAT GAGTGCTAAA GTAACCAACA ATATACAAAA     950
CCTGATAAAG·CAACAATATT GCAATAAGTA CAAGCTTAAA GAAGAAATGG    1000
GTGAGCTCCA TTTTTGCTGG GAGGAGAAAA GCTTAATCGT TGCTTCAGCT    1050
TGGAGGTAAG ATAAGTGACA AGAGCATATA GTCTTTATGT TTCATAAAAC    1100
ATAATTATGT TTTTACTGTA ATCTTGGGTT ATTGTGTAAC TGGTTAAATC    1150
ATCTCCATGT ATTATTACCA GAGGTTAGGG GTGATCACAG GTACTAAAAG    1200
CTAATCTAAC ACTTATGGAA GAATTTTTCT TTCTTTTTTT TCCCTATTAT    1250
ATAAAAATAA TTAGAGTTTT GGTTCTAAAC CTATTTGCTA AGTGTGAATG    1300
AGTCTTTACA TGTTCATATT TCAGTTCAAA TGGTTAAATT TGTTAAGGTT    1350
CTCACTTAAA AAAAAA
```

FIG. 10

Zm-scl1

```
        10         20         30         40         50
CCAGGAGGCGTTCGAGCGGGAGGAGCGTGTGCACATCATCGACCTCGACA
 Q  E  A  F  E  R  E  E  R  V  H  I  I  D  L  D  I

        60         70         80         90        100
TCATGCAGGGGCTGCAGTGGCCGGGCCTCTTCCACATCCTTGCCTCCCGC
 M  Q  G  L  Q  W  P  G  L  F  H  I  L  A  S  R
```

FIG. 11 A

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
CCACGCGTCCG TCAAAGGATA CAACCATGTA CACATAATTG ACTTTTCCCT   50
H  A  S  V  K  G  Y  N  H  V  H  I  I  D  F  S  L

GATGCAAGGT CTCCAGTGGC CGGCACTCAT GGATGTCTTC TCCGCCCGTG  100
M  Q  G  L  Q  W  P  A  L  M  D  V  F  S  A  R  E

AGGGTGGGCC ACCAAAGCTC CGAATCACAG GCATTGGCCC GAACCCAATA  150
G  G  P  P  K  L  R  I  T  G  I  G  P  N  P  I

GGTGGCCGTG ACGAGCTCCA TGAAGTGGGA ATTCGCCTCG CCAAGTATGC  200
G  G  R  D  E  L  H  E  V  G  I  R  L  A  K  Y  A

ACACTCGGTG GGTATCGACT TCACTTTCCA GGGAGTCTGT GTCGATCAAC  250
H  S  V  G  I  D  F  T  F  Q  G  V  C  V  D  Q  L

TTGATAGGTT GTCCGACTGG ATGCTTCTCA AACCAATCAA AGGAGAGGCA  300
D  R  L  C  D  W  M  L  L  K  P  I  K  G  E  A

GTTGCCATAA ACTCCATCCT ACAACTCCAT CGCCTCCTCG TTGACCCAGA  350
V  A  I  N  S  I  L  Q  L  H  R  L  L  V  D  P  D

TGCAAACCCA GTCGTCCCCG CACCAATAGA TATCCTCCTC AAATTGGTCA  400
A  N  P  V  V  P  A  P  I  D  I  L  L  K  L  V  I

TCAAGATAAA CCCCATGATC TTCACGGTGG TTGAGCATGA GGCAGATCAC  450
K  I  N  P  M  I  F  T  V  V  E  H  E  A  D  H

AACAGACCAC CACTACTAGA GAGGTTCACT AATGCCCTCT TCCACTATGC  500
N  R  P  P  L  L  E  R  F  T  N  A  L  F  H  Y  A

GACCATGTTT GACTCTTTGG AGGCCATGCA TCGTTGTACC AGTGGTAGAG  550
T  M  F  D  S  L  E  A  M  H  R  C  T  S  G  R  D

ACATCACCGA CTCACTCACA GAGGTGTACC TTCGAGGTGA GATTTTTGAC  600
I  T  D  S  L  T  E  V  Y  L  R  G  E  I  F  D
```

FIG. 11B

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
ATTGTCTGCG GCGAGGGCAG TGCACGCACC GAACGTCATG AGTTGTTTGG   650
I  V  C  G   E  G  S   A  R  T   E  R  H  E   L  F  G

TCACTGGAGG GAGAGGCTCA CCTATGCTGG GCTAACTCAA GTGTGGTTCG   700
 H  W  R   E  R  L  T   Y  A  G   L  T  Q   V  W  F  D

ACCCCGATGA GGTTGACACG CTAAAAGACC AGTTGATCCA TGTGACATCC   750
  P  D  E   V  D  T   L  K  D  Q   L  I  H   V  T  S

TTATCTGGCT CTGGGTTCAA CATCCTAGTG TGTGATGGCA GCCTTGCACT   800
L  S  G  S   G  F  N   I  L  V   C  D  G  S   L  A  L

AGCGTGGCAT AATCGCCCGT TATATGTGGC AACAGCTTGG TGTGTGACAG   850
 A  W  H   N  R  P  L   Y  V  A   T  A  W   C  V  T  G

GAGGAAATGC TGCCAGTTCC ATGGTTGGCA ACATCTGTAA GGGTACAAAT   900
  G  N  A   A  S  S   M  V  G  N   I  C  K   G  T  N

GATAGTAGAA GAAAGGAAAA CCGTAATGGA CCCATGGAGT AGCAGGAAGA   950
D  S  R  R   K  E  N   R  N  G   P  M  E  X

ATAACCATGT CATGAGCAAA TCGATCAAGT AATAAAATGC ACTGATGACA  1000

TGCATGGTGA TCTAAAGTTT TTTTGCGTGA ATGTGCAATG ACGAATTGTT  1050

CAATTTGAAT AACCTAATCA TGAGACTCAA AAAAAAAAAA AAA          1093
```

FIG. 11B(cont'd.)

```
CCCAACTTGG GAAGCCCTTC CTCCGCTCCG CCTCCTACCT CAAGGAGGCC      50
CTCCTCCTCG CACTCGCCGA CAGCCACCAT GGCTCCTCCG GCGTCACCTC     100
GCCGCTCGAC GTTGCCCTCA AGCTTGCAGC ATACAAGTCT TTCTCTGACC     150
TGTCACCTGT GCTCCAGTTC ACTAACTTTA CCGCAACAAG GCGCTTCTTG     200
ATGAGATTGG TGGCATGGCA ACTTCCTGCA TCCATGTCAT TGACTTTGAT     250
CTCGGTGTTG GTGGTCAGTG GGCTTCCTTC TTGCAGGAGC TTGCCCACCG     300
CCGGGGAGCT GGAGGTATGG CCTTGCCGTT GTTGAAGCTC ACGGCTTTCA     350
TGTCGACTGC TTCTCACCAT CCACTGGAGC TGCACCTTAC CCAGGATAAC     400
CTCTCTCAGT TTGCCGCAGA GCTCAGAATT CCTTTCGAAT TCAATGCCGT     450
CAGTCTTGAT GCATTCAATC CTGCGGAATC TATTTCTTCC TCTGGTGATG     500
AAGTTGTTGC TGTTAGCCTC CCTGTTGGCT GCTCTGCTCG TGCACCACCG     550
CTGCCAGCGA TTCTTCGGTT GGTGAAACAG CTTTGTCCTA AGGTTGTCGT     600
GGCTATTGAT C
```

FIG. 12A

```
TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TACAGAGCAA CAGCAGTATA      50
ATATTAATTC TGTACCACAC AACCATTTGA TAGGTTAAAT TACCCTCTAG     100
TCTCTACTCA TAAGCAGTGT TTCCAATGAG ATGATCATGG CTAATTGAGC     150
AGAGCATGGC AACAACCTAA AGCAACATCA TTAGCTATAG AGACTGACAC     200
CAATATTCCT AAATCCACTA GGCTAGCTAA TAAGCTGCAA CGAAAAGCAA     250
TATGAAGAGT TCAACAGCTC AAGACAACAA TTTCATTTGC AACATTTAAT     300
TGCAAGAATA AATGGACATT ACTGGAGTGG TCGATGCTTG CAAACGGTGG     350
TGGAACCTTG GTGGAGTGAA GCTTATGGCT GATCAGCACC GCCAAGATGA     400
TATGGATACA AGCTCCCCAC GCTGCCAGTA GAGCGTAAGA GCAGCTCCGC     450
GTTTCTCCAC ATGGAATCCT CGGACCTGCA CCCGCTTCAG GAGGCAGTCT     500
GC
```

FIG. 12B

FIG. 13A

```
SCR     MAESGDFNGGQPPPHSPLRTTSSGSSSSNNRGPPPPPPPPLVMVRKR----LASEMSS
TF1     MKRD---HHQFQGRLSNHGTSSSSSSISKDK--MMMVKKEEDGGGNMDDELLAV----
TF4     MKRDHHHHHQ-------------------DKKTMMM--NEEDDGNGM-DELLAV----

           |-------------- MOTIF I --------------|
SCR     NPDYNNSSRPPRRVSHLLDSNYNTVTPQQPPSLTAAATVSSQPNPPLSVCGFSG
TF1     -LGYKVRSSEMAEVALKLEQLETMMSNAQEDGLSHLATDAAHYNPSELYS----
TF4     -LGYKVRSSEMADVAQKLEQLEVMMSNVQEDDLSQLATETVHYNPAELYT----

SCR     LPVFPSDRGGRNVMMSVQPMDQDSSSSSASPTVWVDAIIRDLIHS----STSVSIPQL
TF1     ----------------------------------WLDNMLSELNPPPLPASSNGLDPVL
TF4     ----------------------------------WLDSMLTDLNPP----SSN-AEYDL

SCR     IQNVRDIIFPCNPNLGALLEYRLRSLMLLDPSSSSDPSPQTFEPLYQISNNPSP
TF1     PSPEICGFPXSDYDLKVIPXNAIYQFPAIDSSSSSNN--Q--------------
TF4     -----KAI-P---------GDILNQF-AIDSASSSN---Q--------------
```

FIG. 13B

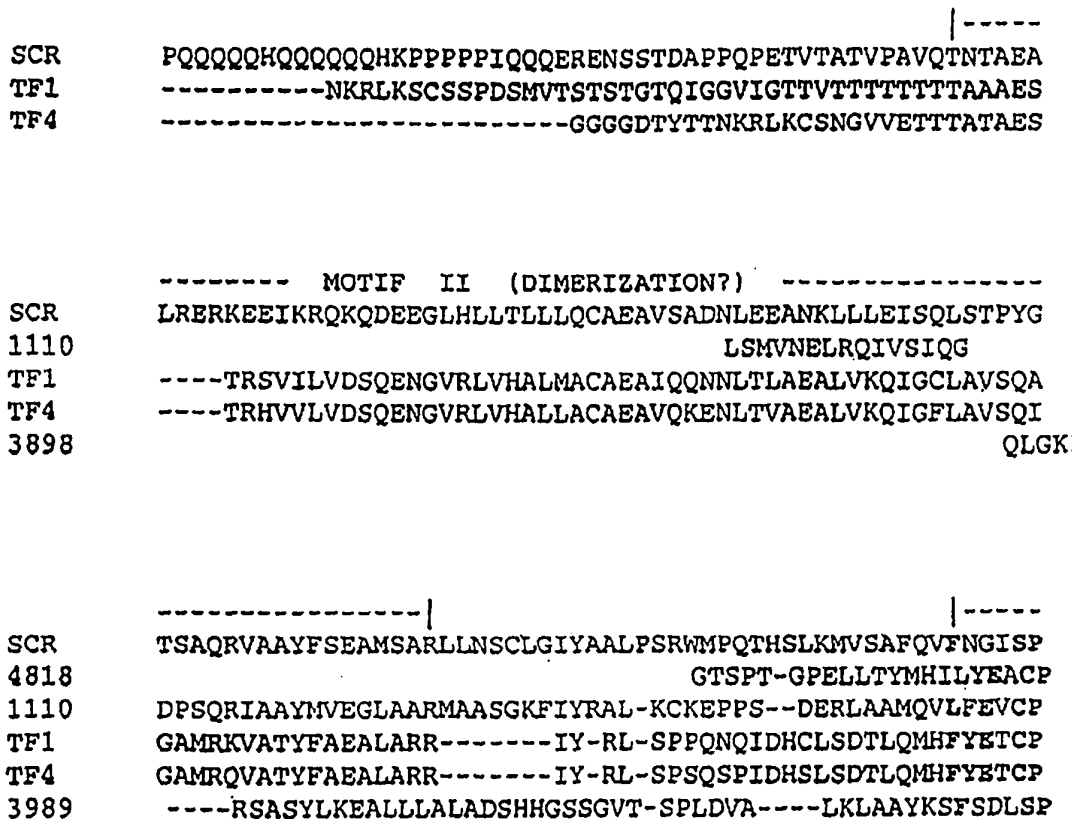

SCR PQQQQQHQQQQQQHKPPPPPIQQQERENSSTDAPPQPETVTATVPAVQTNTAEA
TF1 ----------NKRLKSCSSPDSMVTSTSTGTQIGGVIGTTVTTTTTTTTAAAES
TF4 ----------GGGGDTYTTNKRLKCSNGVVETTTATAES
-------- MOTIF II (DIMERIZATION?) ----------
SCR LRERKEEIKRQKQDEEGLHLLTLLLQCAEAVSADNLEEANKLLLEISQLSTPYG
1110 LSMVNELRQIVSIQG
TF1 ----TRSVILVDSQENGVRLVHALMACAEAIQQNNLTLAEALVKQIGCLAVSQA
TF4 ----TRHVVLVDSQENGVRLVHALLACAEAVQKENLTVAEALVKQIGFLAVSQI
3898 QLGKPFL
SCR TSAQRVAAYFSEAMSARLLNSCLGIYAALPSRWMPQTHSLKMVSAFQVFNGISP
4818 GTSPT-GPELLTYMHILYEACP
1110 DPSQRIAAYMVEGLAARMAASGKFIYRAL-KCKEPPS--DERLAAMQVLFEVCP
TF1 GAMRKVATYFAEALARR-------IY-RL-SPPQNQIDHCLSDTLQMHFYETCP
TF4 GAMRQVATYFAEALARR-------IY-RL-SPSQSPIDHSLSDTLQMHFYETCP
3989 ----RSASYLKEALLLALADSHHGSSGVT-SPLDVA----LKLAAYKSFSDLSP

FIG. 13C

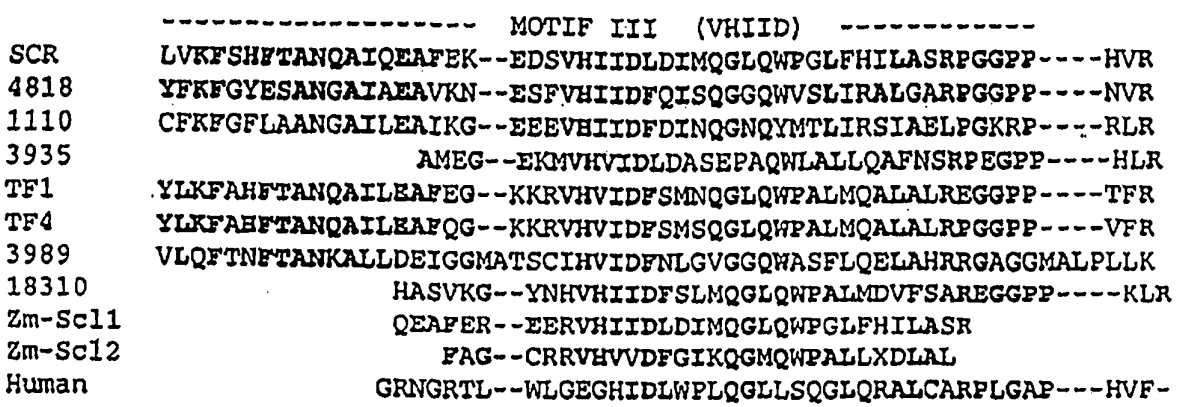

---------- MOTIF III (VHIID) ----------
SCR LVKFSHFTANQAIQEAFEK--EDSVHIIDLDIMQGLQWPGLFHILASRPGGPP----HVR
4818 YFKFGYESANGAIAEAVKN--ESFVHIIDFQISQGGQWVSLIRALGARPGGPP----NVR
1110 CFKFGFLAANGAILEAIKG--EEEVHIIDFDINQGNQYMTLIRSIAELPGKRP---RLR
3935 AMEG--EKMVHVIDLDASEPAQWLALLQAFNSRPEGPP----HLR
TF1 YLKFAHFTANQAILEAFEG--KKRVHVIDFSMNQGLQWPALMQALALREGGPP----TFR
TF4 YLKFAHFTANQAILEAFQG--KKRVHVIDFSMSQGLQWPALMQALALRPGGPP----VFR
3989 VLQFTNFTANKALLDEIGGMATSCIHVIDFNLGVGGQWASFLQELAHRRGAGGMALPLLK
18310 HASVKG--YNHVHIIDFSLMQGLQWPALMDVFSAREGGPP----KLR
Zm-Scl1 QEAFER--EERVHIIDLDIMQGLQWPGLFHILASR
Zm-Scl2 FAG--CRRVHVVDFGIKQGMQWPALLXDLAL
Human GRNGRTL--WLGEGHIDLWPLQGLLSQGLQRALCARPLGAP---HVF-

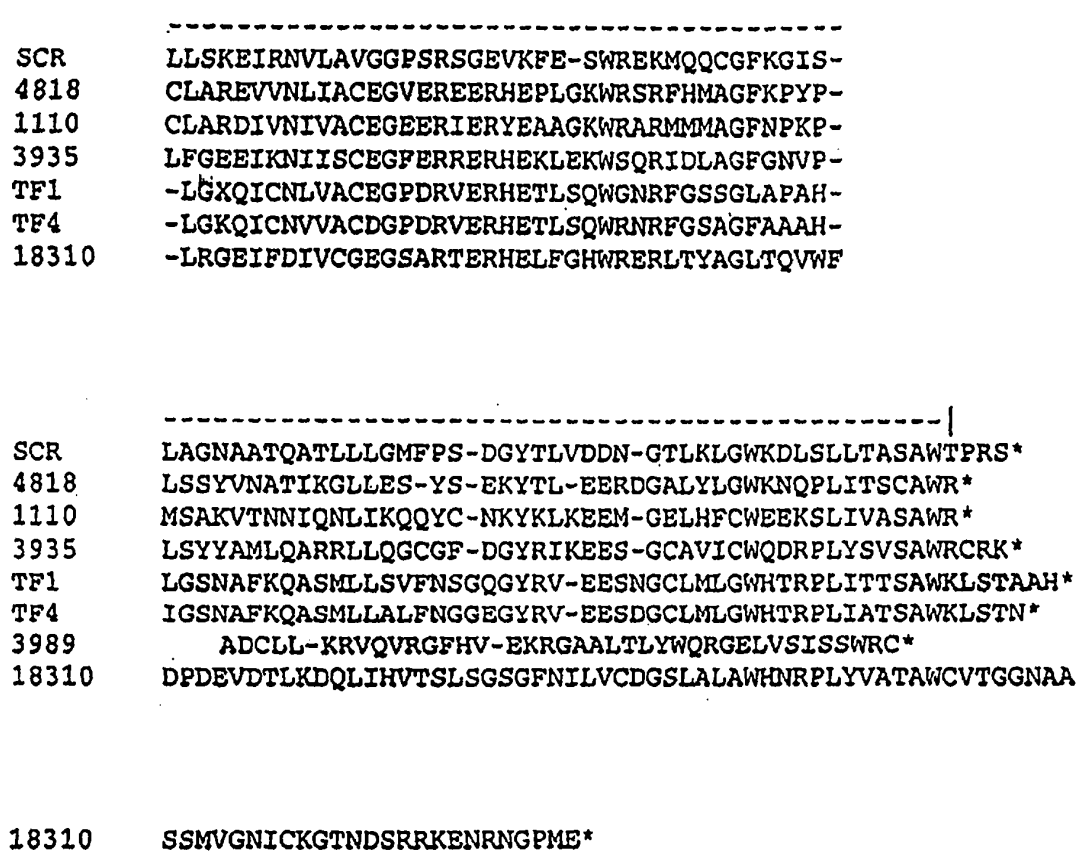
FIG. 13B

```
                                                    |-------------- MOTIF  V  -
          --||-------------- MOTIF  IV  (DIMERIZATION ) ------------------|
SCR       LTG  LGTSMEA       LQATGKR  LSDFTDK  LGLPFEFCPLAEKVGNDLTERLNV
4818      ITGIDDPRSSFARQGG   LELVGQR  LGKLAEM  CGVPFEFHGAALCCTEVEIEKLGV
1110      LTGIDDPESVQRSIGG   LRIINLR  LEQLAED  NGVSFKFKAMPSKTSIVSPSTLGC
3935      ITG  VHHQKEV       LEQMAHR  LIEEAEK  LDIPFQFNPVVSRLDCLNVEQLRV
TF1       LTGIGPPAPDNSDH     LHEVGCK  LAQLAEA  IHVEFEYRGF  VANSLAD  LDASMLELRP
TF4       LTGIGPPAPDNFDY     LHEVGCK  LAHLAEA  IHVEFEYRGF  VANTLAD  LDASMLELRP
3989      LTAFMSTASHHPLE      LHLTQDN  LSQFAAE  LRIPFEFNAVSLDAFNPAESISSSGDE
18310     ITGIGPNPIGGRDE      LHEVGIR  LAKYAHS  VGIDFTFQGVCVDQLDRLCDWMLLKPI
Human     LPGLHTLS...        LGLQXRH  LLVHMMA  LSYSYGRXP...
```

```
          ---------------|
SCR       RKREAAVHWLQHSLYDVTGSDAHTLWLL---QRLAPK--------------------
4818      RNGEALAVNFPLVLHHMPDESVTVENHR---DRLLRL--------------------
1110      KPGETL VNFAFQLHHMPDESVTTVNQR---DELLHM--------------------
3935      KTGEALAVSSVLQLHTFLASDDDLMRKNC-ALRFQNNPSGVDLQRVLMMSHGS
TF1       SDTEAVAVNSVFELHKLLGRXGGIEKVLG-------------------------
TF4       SEIESVAVNSVFELHKLLGRPGAIDKVLG-------------------------
18310     K-GEAVAINSILQLHRLLVDPDANPVVPAPIDILLK---
3989      VVAVSLPVGCSARAPPLPAILRLVKQLCPKVVVAID
```

```
                                                      |----
SCR      -----------------------------------------------VVTV-
4818     ---------------------------------------VKHLSPN-VVTL-
1110     ---------------------------------------VKSLNPK-LVTV-
3935     AAEARENDMSNNNGYSPSGDSASSLPLPSSGRTDSFLNAIWGLSPKVMVVT-
TF1      ---------------------------------------VVKQD*TGDFHXW
TF4      ---------------------------------------VVNQIKPEIFTV-
18310    ---------------------------------------LVIKINPMIFTV-

         ----------------- MOTIF  VI ---------------------
SCR      VEQDLSHAGS--FLG-RFVEAIHYYSALFDSLGASYGEESE---ERHVVEQQ
4818     VEQEANTNTAP-FLP-RFVETMNHYLAVFESIDVKLARDHK---ERINVEQH
1110     VEQDVNTNTSP-FFP-RFIEAYEYYSAVFESLDMTLPRESQ---ERMNVERQ
3935     -EQDSDHNGS--TLMERLLESLYTYAALFDCLETKVPRTSQ---DRIKVEKM
TF1      XRQEPNHNG-PGFLD-GXTESLHYYSTXFDSLEG--XPNSQ---DKLMSEXY
TF4      VEQESNHNS-PIFLD-RFTESLHYYSTLFDSLEG--VPSGQ---DKVMSEVY
18310    VEHEADHNR-PPLLE-RFTNALFHYATMFDSLEAMHTCTSGRDITDSLTEVY
```

FIG. 13F


SCR LLSKEIRNVLAVGGPSRSGEVKFE-SWREKMQQCGFKGIS-
4818 CLAREVVNLIACEGVEREERHEPLGKWRSRFHMAGFKPYP-
1110 CLARDIVNIVACEGEERIERYEAAGKWRARMMMAGFNPKP-
3935 LFGEEIKNIISCEGFERRERHEKLEKWSQRIDLAGFGNVP-
TF1 -LGXQICNLVACEGPDRVERHETLSQWGNRFGSSGLAPAH-
TF4 -LGKQICNVVACDGPDRVERHETLSQWRNRFGSAGFAAAH-
18310 -LRGEIFDIVCGEGSARTERHELFGHWRERLTYAGLTQVWF
SCR LAGNAATQATLLLGMFPS-DGYTLVDDN-GTLKLGWKDLSLLTASAWTPRS*
4818 LSSYVNATIKGLLES-YS-EKYTL-EERDGALYLGWKNQPLITSCAWR*
1110 MSAKVTNNIQNLIKQQYC-NKYKLKEEM-GELHFCWEEKSLIVASAWR*
3935 LSYYAMLQARRLLQGCGF-DGYRIKEES-GCAVICWQDRPLYSVSAWRCRK*
TF1 LGSNAFKQASMLLSVFNSGQGYRV-EESNGCLMLGWHTRPLITTSAWKLSTAAH*
TF4 IGSNAFKQASMLLALFNGGEGYRV-EESDGCLMLGWHTRPLIATSAWKLSTN*
3989 ADCLL-KRVQVRGFHV-EKRGAALTLYWQRGELVSISSWRC*
18310 DPDEVDTLKDQLIHVTSLSGSGFNILVCDGSLALAWHNRPLYVATAWCVTGGNAA
18310 SSMVGNICKGTNDSRRKENRNGPME*

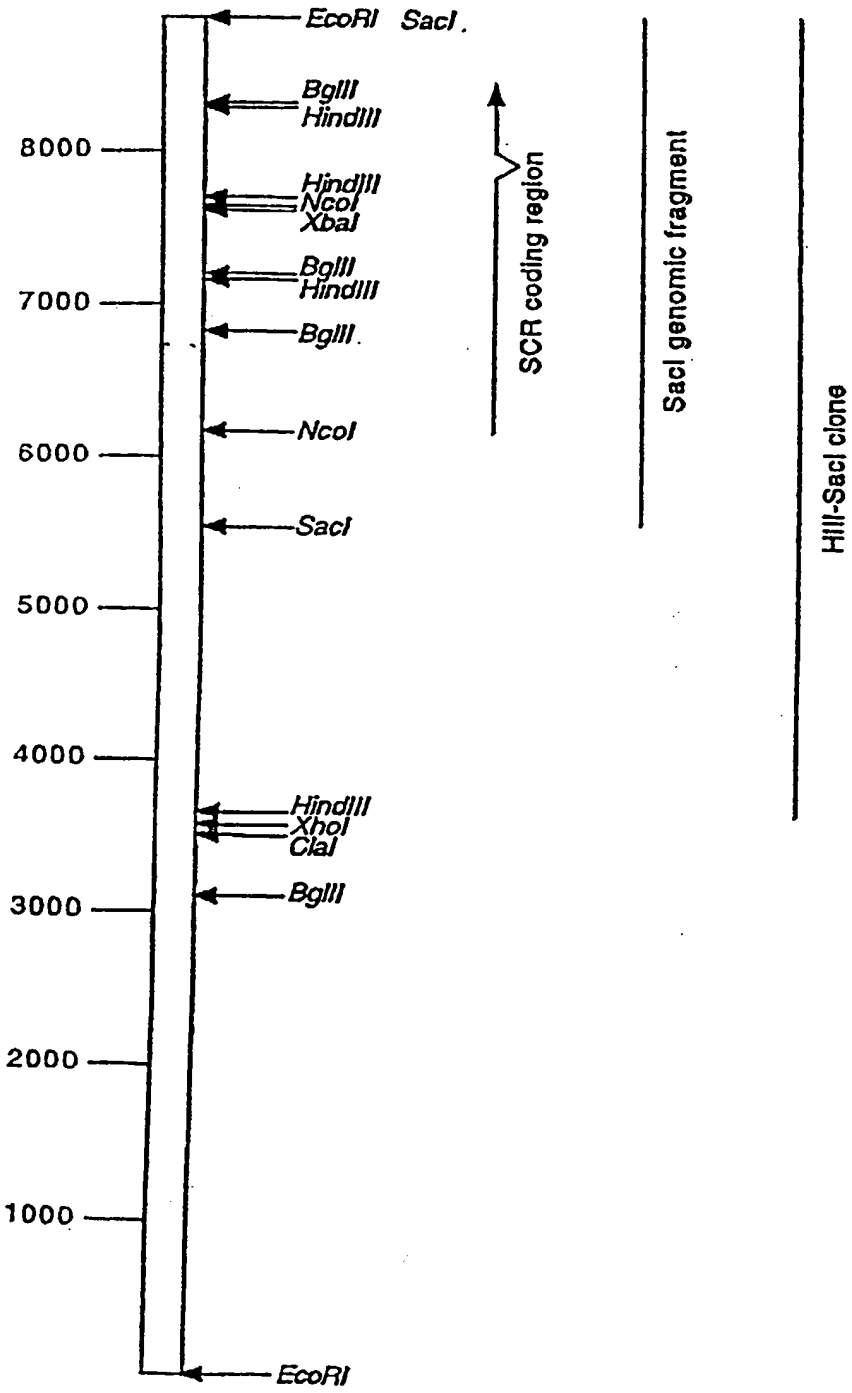

EcoRI
SacI
BglII
HindIII
HindIII
NcoI
XbaI
BglII
HindIII
BglII
NcoI
SacI
HindIII
XhoI
ClaI
BglII
EcoRI
8000
7000
6000
5000
4000
3000
2000
1000
SCR coding region
SacI genomic fragment
HIII-SacI clone

FIG. 14

FIG. 15A

Old Name						New Name
Scr	.....	.....	.....	.....	.....	*SCR*
3989	.....	.....	.....	.....	.....	*SRPo3*
12398	.....	.....	.....	.....	.....	*SRPa6*
4871	.....	.....	.....	.....	.....	*SRPa5*
11846	.....	.....	.....	.....	.....	*SRPo4*
2504	.....	.....	.....	.....	.....	*SRPo2*
3935	.....	.....	.....	.....	.....	*SRPa3*
11261	.....	.....	.....	.....	.....	*SRPa10*
713	.....	.....	.....	.....	.....	*SRPo1*
10964	.....	.....	.....	.....	.....	*SRPa9*
23196	LL	KVLLCHLVAE	STKRRIKIRP	LLDINDSGFL	GFWSWIHMGS	*SRPa12*
Tf1	.....	.....	.....	.....	.....	*SRPa8*
Tf4	.....	.....	.....	.....	.....	*SRPa2*
18310	.....	.....	.....	.....	.....	*SRPm1*
18652	.....	.....	.....	.....	.....	*SRPa11*
4818	.....	.....	.....	.....	.....	*SRPa4*
21729	.....	.....	.....	.....	.....	*SRPa7*
1110	.....	.....	.....	.....	.....	*SRPa1*
174	.....	.....	.....	.....	.....	*SRPb1*
33/08	.....	.....	.....	.....	.....	*SRPa13*
	-150				-101	

FIG. 15B

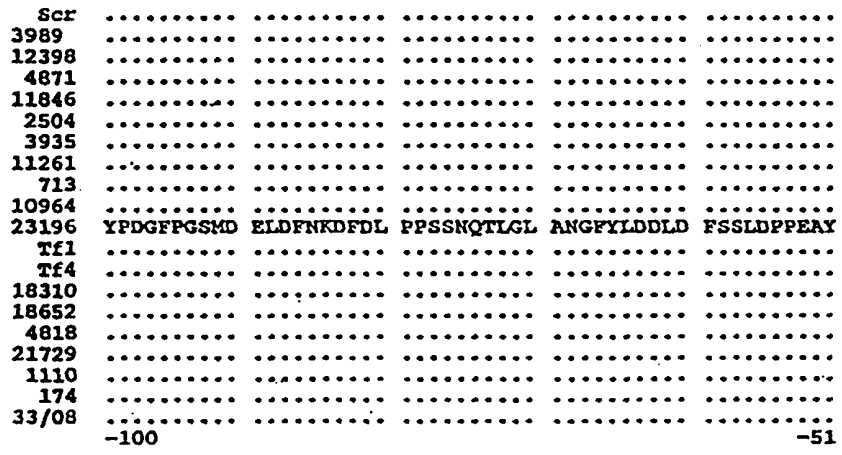

Scr
3989
12398
4871
11846
2504
3935
11261
713
10964
23196 YPDGFPGSMD ELDFNKDFDL PPSSNQTLGL ANGFYLDDLD FSSLDPPEAY
Tf1
Tf4
18310
18652
4818
21729
1110
174
33/08
-100
-51

FIG. 15C

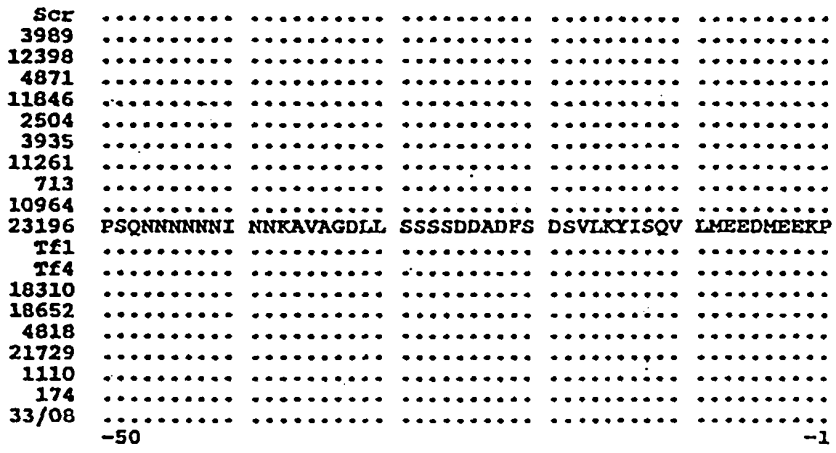

Scr
3989
12398
4871
11846
2504
3935
11261
713
10964
23196
Tf1
Tf4
18310
18652
4818
21729
1110
174
33/08
PSQNNNNNNI NNKAVAGDLL SSSSDDADFS DSVLKYISQV LMEEDMEEKP
-50
-1

FIG. 15D

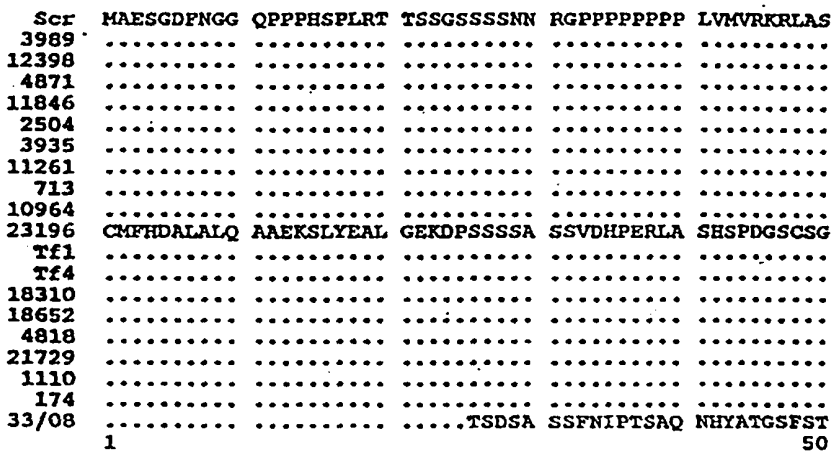

Scr MAESGDPNGG QPPPHSPLRT TSSGSSSSNN RGPPPPPPPP LVMVRKRLAS
3989 ....
12398 ....
4871 ....
11846 ....
2504 ....
3935 ....
11261 ....
713 ....
10964 ....
23196 CMFHDALALQ AAEKSLYEAL GEKDPSSSSA SSVDHPERLA SHSPDGSCSG
Tf1 ....
Tf4 ....
18310 ....
18652 ....
4818 ....
21729 ....
1110 ....
174 ....
33/08 ....TSDSA SSFNIPTSAQ NHYATGSFST
1
50

FIG. 15E

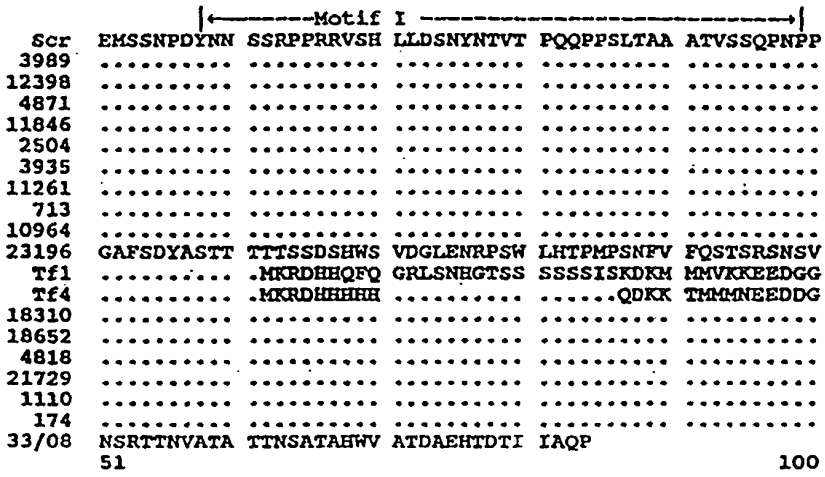

|←Motif I →|
Scr EMSSNPDYNN SSRPPRRVSH LLDSNYNTVT PQQPPSLTAA ATVSSQPNPP
3989 ....
12398 ....
4871 ....
11846 ....
2504 ....
3935 ....
11261 ....
713 ....
10964 ....
23196 GAFSDYASTT TTTSSDSHWS VDGLENRPSW LHTPMPSNFV FQSTSRSNSV
Tf1 .... .MKRDHHQFQ GRLSNHGTSS SSSSISKDKM MMVKKEEDGG
Tf4 .... .MKRDHHHHH ....QDKK TMMMNEEDDG
18310 ....
18652 ....
4818 ....
21729 ....
1110 ....
174 ....
33/08 NSRTTNVATA TTNSATAHWV ATDAEHTDTI IAQP
51 100

FIG. 15F

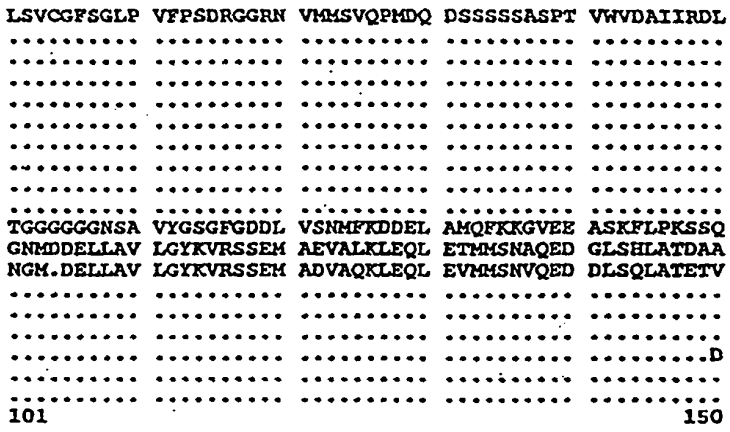

Scr LSVCGFSGLP VFPSDRGGRN VMMSVQPMDQ DSSSSSASPT VWVDAIIRDL
3989 ....
12398 ....
4871 ....
11846 ....
2504 ....
3935 ....
11261 ....
713 ....
10964 ....
23196 TGGGGGGNSA VYGSGFGDDL VSNMFKDDEL AMQFKKGVEE ASKFLPKSSQ
Tf1 GNMDDELLAV LGYKVRSSEM AEVALKLEQL ETMMSNAQED GLSHLATDAA
Tf4 NGM.DELLAV LGYKVRSSEM ADVAQKLEQL EVMMSNVQED DLSQLATETV
18310 ....
18652 ....
4818 ....
21729 ....D
1110 ....
174 ....
101
150

FIG. 15G

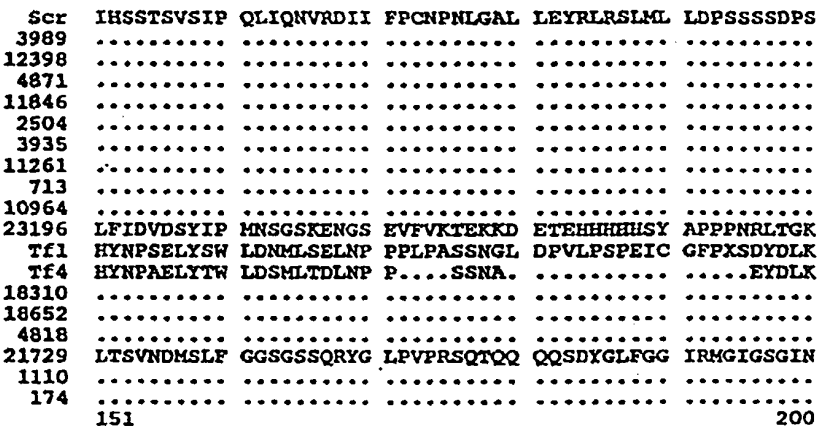

Scr IHSSTSVSIP QLIQNVRDII FPCNPNLGAL LEYRLRSLML LDPSSSSDPS
3989 ....
12398 ....
4871 ....
11846 ....
2504 ....
3935 ....
11261 ....
713 ....
10964 ....
23196 LFIDVDSYIP MNSGSKENGS EVFVKTEKKD ETEHHHHHSY APPPNRLTGK
Tf1 HYNPSELYSW LDNMLSELNP PPLPASSNGL DPVLPSPEIC GFPXSDYDLK
Tf4 HYNPAELYTW LDSMLTDLNP P....SSNA. ....EYDLK
18310 ....
18652 ....
4818 ....
21729 LTSVNDMSLF GGSGSSQRYG LPVPRSQTQQ QQSDYGLFGG IRMGIGSGIN
1110 ....
174 ....
151
200

FIG. 15H

```
  Scr  PQTFEPLYQI SNNPSPPQQQ QQHQQQQQQH KPPPPPIQQQ ERENSSTDAP
 3989  .......... .......... .......... .......... ..........
12398  .......... .......... .......... .......... ..........
 4871  .......... .......... .......... .......... ..........
11846  .......... .......... .......... .......... ..........
 2504  .......... .......... .......... .......... ..........
 3935  .......... .......... .......... .......... ..........
11261  .......... .......... .......... .......... ..........
  713  .......... .......... .......... .......... ..........
10964  .......... .......... .......... .......... ..........
23196  KSHWRDEDED VEERSNKQSA VYVEESELSE MFDNMFLCGP GKPVCILNQN
  Tf1  VIPXNAIYQP PAIDSSSSSN NQ........ NKRLKSCSSP DSMVTSTSTG
  Tf4  AIPGDAILNQ FAIDSASSSN QGGGGDTYTT NKRLKCS... ..........
18310  .......... .......... .......... .......... ..........
18652  .......... .......... .......... .......... ..........
 4818  .......... .......... .......... .......... ..........
21729  NYPTLTGVPC IEPVQNRVHE SENMLNSLRE LEKQLLDDDD ESGGDDDVSV
 1110  .......... .......... .......... .......... ..........
  174  .......... .......... .......... .......... ..........
       201                                                 250
```

FIG. 15I

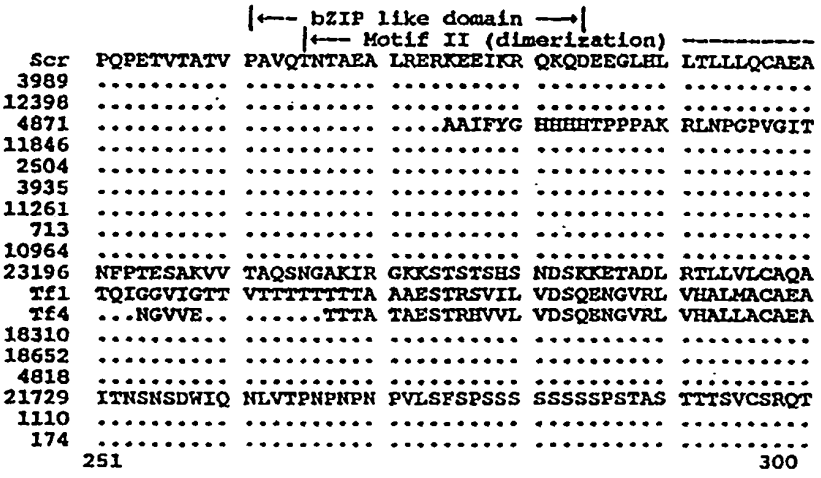

|←— bZIP like domain —→|
|←— Motif II (dimerization) ——————
Scr PQPETVTATV PAVQTNTAEA LRERKEEIKR QKQDEEGLHL LTLLLQCAEA
3989 ....
12398 ....
4871 ....AAIFYG HHHHTPPPAK RLNPGPVGIT
11846 ....
2504 ....
3935 ....
11261 ....
713 ....
10964 ....
23196 NFPTESAKVV TAQSNGAKIR GKKSTSTSHS NDSKKETADL RTLLVLCAQA
Tf1 TQIGGVIGTT VTTTTTTTTA AAESTRSVIL VDSQENGVRL VHALMACAEA
Tf4 ...NGVVE.. ......TTTA TAESTRHVVL VDSQENGVRL VHALLACAEA
18310 ....
18652 ....
4818 ....
21729 ITNSNSDWIQ NLVTPNPNPN PVLSFSPSSS SSSSSPSTAS TTTSVCSRQT
1110 ....
174 ....
251
300

FIG. 15J

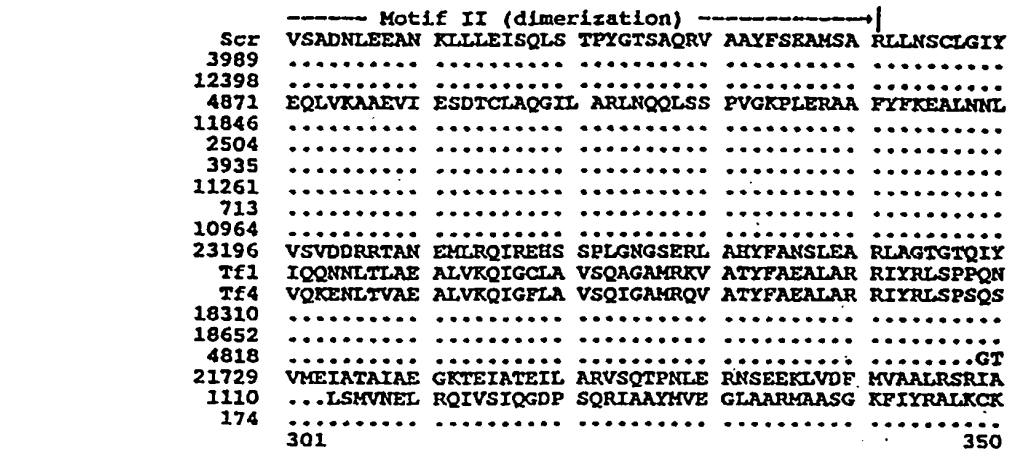

------ Motif II (dimerization) ----------→|
Scr VSADNLEEAN KLLLEISQLS TPYGTSAQRV AAYFSRAMSA RLLNSCLGIY
3989 ....
12398 ....
4871 EQLVKAAEVI ESDTCLAQGIL ARLNQQLSS PVGKPLERAA FYFKEALNNL
11846 ....
2504 ....
3935 ....
11261 ....
713 ....
10964 ....
23196 VSVDDRRTAN EMLRQIREHS SPLGNGSERL AHYFANSLEA RLAGTGTQIY
Tf1 IQQNNLTLAE ALVKQIGCLA VSQAGAMRKV ATYFAEALAR RIYRLSPPQN
Tf4 VQKENLTVAE ALVKQIGFLA VSQIGAMRQV ATYFAEALAR RIYRLSPSQS
18310 ....
18652 ....
4818 ....GT
21729 VMEIATAIAE GKTEIATEIL ARVSQTPNLE RNSEEKLVDF MVAALRSRIA
1110 ...LSMVNEL RQIVSIQGDP SQRIAAYHVE GLAARMAASG KFIYRALKCK
174 ....
301 350

FIG. 15K

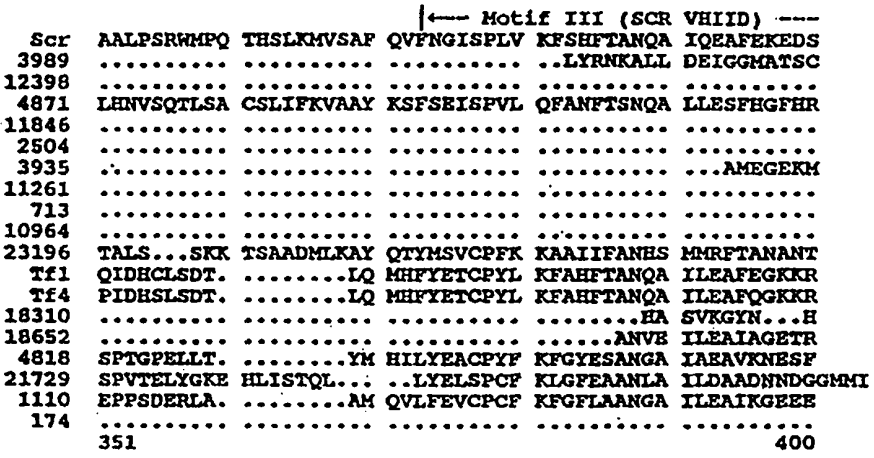

|<— Motif III (SCR VHIID) —
Scr AALPSRWMPQ THSLKMVSAF QVFNGISPLV KFSHFTANQA IQEAFEKEDS
3989 ....LYRNKALL DEIGGMATSC
12398 ....
4871 LHNVSQTLSA CSLIFKVAAY KSFSEISPVL QFANFTSNQA LLESFHGFHR
11846 ....
2504 ....
3935 ....AMEGEKM
11261 ....
713 ....
10964 ....
23196 TALS...SKK TSAADMLKAY QTYMSVCPFK KAAIIFANHS MMRFTANANT
Tf1 QIDHCLSDT. .....LQ MHFYETCPYL KFAHFTANQA ILEAFEGKKR
Tf4 PIDHSLSDT. .....LQ MHFYETCPYL KFAHFTANQA ILEAFQGKKR
18310 ....HA SVKGYN...H
18652 ....ANVE ILEAIAGETR
4818 SPTGPELLT. .....YM HILYEACPYF KFGYESANGA IAEAVKNESF
21729 SPVTELYGKE HLISTQL... ..LYELSPCF KLGFEAANLA ILDAADNNDGGMMI
1110 EPPSDERLA. .....AM QVLFEVCPCF KFGFLAANGA ILEAIKGEEE
174 ....
351
400

FIG. 15L

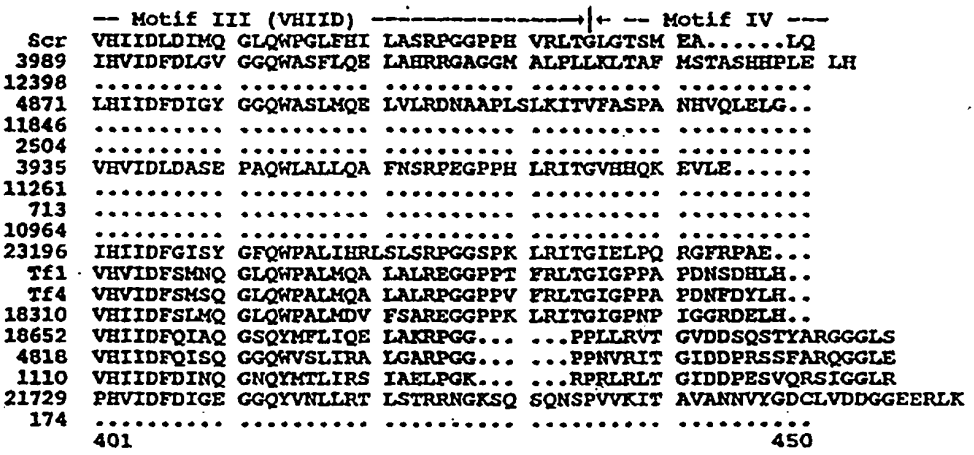

-- Motif III (VHIID) ----------→|← -- Motif IV ---
Scr VHIIDLDIMQ GLQWPGLFHI LASRPGGPPH VRLTGLGTSM EA.....LQ
3989 IHVIDFDLGV GGQWASFLQE LAHRRGAGGM ALPLLKLTAF MSTASHHPLE LH
12398 .....
4871 LHIIDFDIGY GGQWASLMQE LVLRDNAAPLSLKITVFASPA NHVQLELG..
11846 .....
2504 .....
3935 VHVIDLDASE PAQWLALLQA FNSRPEGPPH LRITGVHHQK EVLE.....
11261 .....
713 .....
10964 .....
23196 IHIIDFGISY GFQWPALIHRLSLSRPGGSPK LRITGIELPQ RGFRPAE...
Tf1 VHVIDFSMNQ GLQWPALMQA LALREGGPPT FRLTGIGPPA PDNSDHLH..
Tf4 VHVIDFSMSQ GLQWPALMQA LALRPGGPPV FRLTGIGPPA PDNFDYLH..
18310 VHIIDFSLMQ GLQWPALMDV FSAREGGPPK LRITGIGPNP IGGRDELH..
18652 VHIIDFQIAQ GSQYMFLIQE LAKRPGG... ...PPLLRVT GVDDSQSTYARGGGLS
4818 VHIIDFQISQ GGQWVSLIRA LGARPGG... ...PPNVRIT GIDDPRSSFARQGGLE
1110 VHIIDFDINQ GNQYMTLIRS IAELPGK... ...RPRLRLT GIDDPESVQRSIGGLR
21729 PHVIDFDIGE GGQYVNLLRT LSTRRNGKSQ SQNSPVVKIT AVANNVYGDCLVDDGGEERLK
174 .....
401 450

FIG. 15M

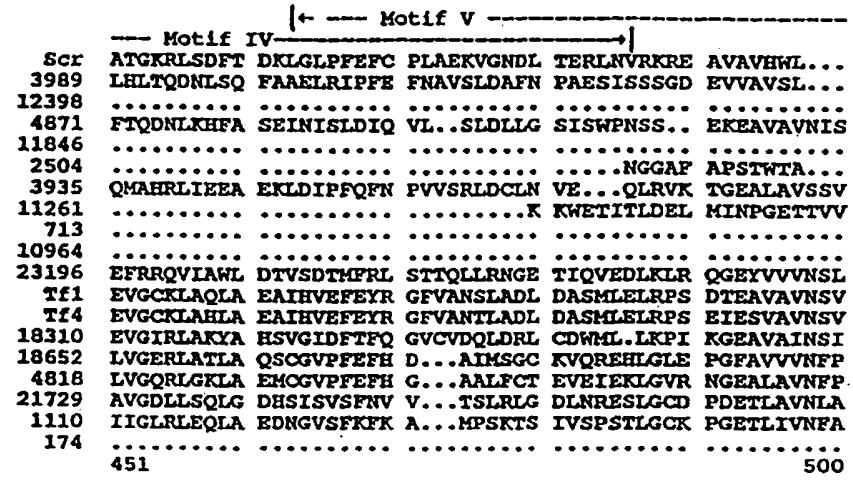

|← --- Motif V ----------
--- Motif IV----------→|
Scr ATGKRLSDFT DKLGLPFEFC PLAEKVGNDL TERLNVRKRE AVAVHWL...
3989 LHLTQDNLSQ FAAELRIPFE FNAVSLDAFN PAESISSSGD EVVAVSL...
12398 ....
4871 FTQDNLKHFA SEINISLDIQ VL..SLDLLG SISWPNSS.. EKEAVAVNIS
11846 ....
2504 ....NGGAF APSTWTA...
3935 QMAHRLIEEA EKLDIPFQFN PVVSRLDCLN VE...QLRVK TGEALAVSSV
11261 ....K KWETITLDEL MINPGETTVV
713 ....
10964 ....
23196 EFRRQVIAWL DTVSDTMFRL STTQLLRNGE TIQVEDLKLR QGEYVVVNSL
Tf1 EVGCKLAQLA EAIHVEFEYR GFVANSLADL DASMLELRPS DTEAVAVNSV
Tf4 EVGCKLAHLA EAIHVEFEYR GFVANTLADL DASMLELRPS EIESVAVNSV
18310 EVGIRLAKYA HSVGIDFTFQ GVCVDQLDRL CDWML.LKPI KGEAVAINSI
18652 LVGERLATLA QSCGVPFEFH D...AIMSGC KVQREHLGLE PGFAVVVNFP
4818 LVGQRLGKLA EMCGVPFEFH G...AALFCT EVEIEKLGVR NGEALAVNFP
21729 AVGDLLSQLG DHSISVSFNV V...TSLRLG DLNRESLGCD PDETLAVNLA
1110 IIGLRLEQLA EDNGVSFKFK A...MPSKTS IVSPSTLGCK PGETLIVNFA
174 ....
451 500

FIG. 15N

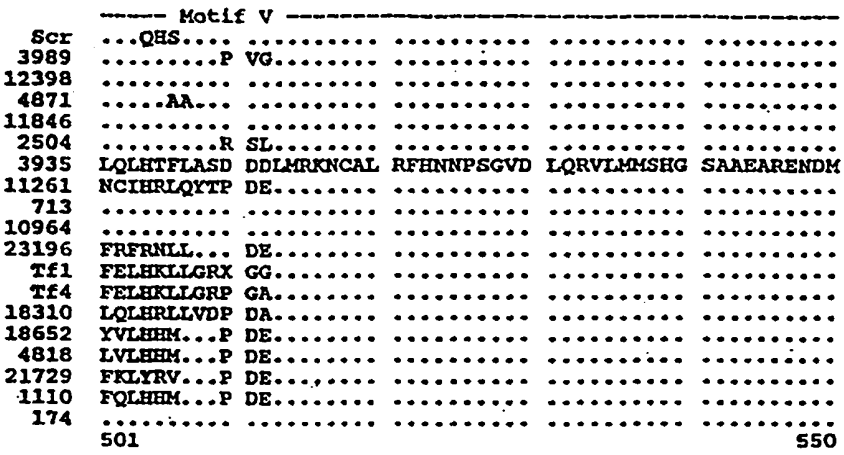

------ Motif V ----------
Scr ...QHS.... ....
3989 ....P VG........ ....
12398 ....
4871 .....AA... ....
11846 ....
2504 ....R SL........ ....
3935 LQLHTFLASD DDLMRKNCAL RFHNNPSGVD LQRVLMMSHG SAAEARENDM
11261 NCIHRLQYTP DE........ ....
713 ....
10964 ....
23196 FRFRNLL... DE........ ....
Tf1 FELHKLLGRX GG........ ....
Tf4 FELHKLLGRP GA........ ....
18310 LQLHRLLVDP DA........ ....
18652 YVLHHM...P DE........ ....
4818 LVLHHM...P DE........ ....
21729 FKLYRV...P DE........ ....
1110 FQLHHM...P DE........ ....
174 ....
501
550

FIG. 150

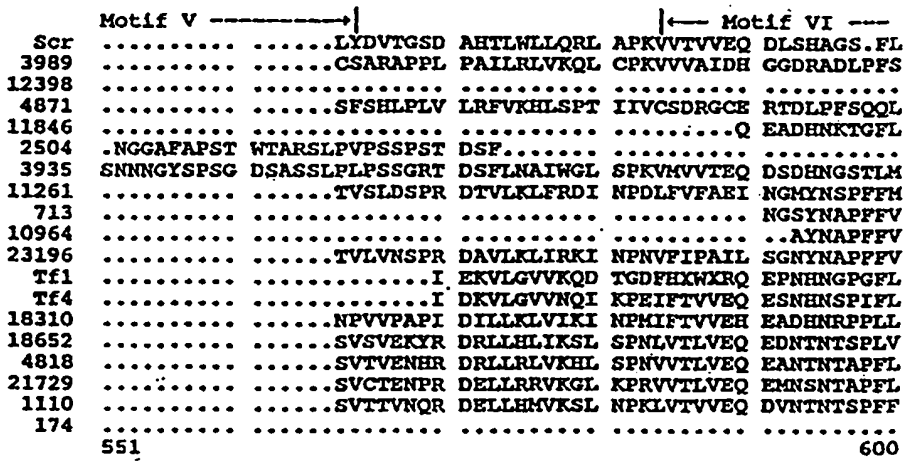

Motif V
Motif VI
Scr ....LYDVTGSD AHTLWLLQRL APKVVTVVEQ DLSHAGS.FL
3989 ....CSARAPPL PAILRLVKQL CPKVVVAIDH GGDRADLPFS
12398 ....
4871 ....SFSHLPLV LRFVKHLSPT IIVCSDRGCE RTDLPFSQQL
11846 ....Q EADHNKTGFL
2504 .NGGAFAPST WTARSLPVPSSPST DSF..... ....
3935 SNNNGYSPSG DSASSLPLPSSGRT DSFLNAIWGL SPKVMVVTEQ DSDHNGSTLM
11261 ....TVSLDSPR DTVLKLFRDI NPDLFVFAEI NGMYNSPFFM
713 .... NGSYNAPFFV
10964 ....AYNAPFFV
23196 ....TVLVNSPR DAVLKLIRKI NPNVFIPAIL SGNYNAPFFV
Tf1 ....I EKVLGVVKQD TGDFHXWXRQ EPNHNGPGFL
Tf4 ....I DKVLGVVNQI KPEIFTVVEQ ESNHNSPIFL
18310 ....NPVVPAPI DILLKLVIKI NPHIFTVVEH EADHNRPPLL
18652 ....SVSVEKYR DRLLHLIKSL SPNLVTLVEQ EDNTNTSPLV
4818 ....SVTVENHR DRLLRLVKHL SPNVVTLVEQ EANTNTAPFL
21729 ....SVCTENPR DELLRRVKGL KPRVVTLVEQ EMNSNTAPFL
1110 ....SVTTVNQR DELLHMVKSL NPKLVTVVEQ DVNTNTSPFF
174 ....
551
600

FIG. 15P

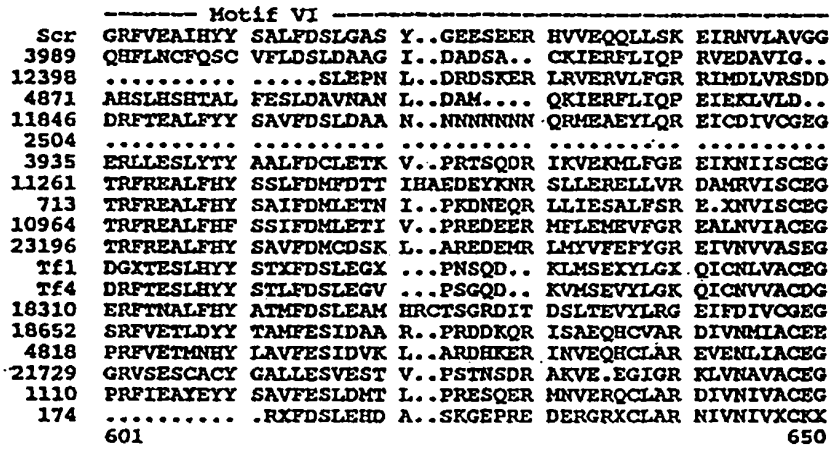

Motif VI
Scr GRFVEAIHYY SALFDSLGAS Y..GEESEER HVVEQQLLSK EIRNVLAVGG
3989 QHFLNCFQSC VFLDSLDAAG I..DADSA.. CKIERFLIQP RVEDAVIG..
12398 ....SLEPN L..DRDSKER LRVERVLFGR RIMDLVRSDD
4871 AHSLHSHTAL FESLDAVNAN L..DAM.... QKIERFLIQP EIEKLVLD..
11846 DRFTEALFYY SAVFDSLDAA N..NNNNNNN QRMEAEYLQR EICDIVCGEG
2504 ....
3935 ERLLESLYTY AALFDCLETK V..PRTSQDR IKVEKMLFGE EIKNIISCEG
11261 TRFREALFHY SSLFDMFDTT IHAEDEYKNR SLLERELLVR DAMRVISCEG
713 TRFREALFHY SAIFDMLETN I..PKDNEQR LLIESALFSR E.XNVISCEG
10964 TRFREALFHF SSIFDMLETI V..PREDEER MFLEMEVFGR EALNVIACEG
23196 TRFREALFHY SAVFDMCDSK L..AREDEMR LMYVFEFYGR EIVNVVASEG
Tf1 DGXTESLHYY STXFDSLEGX ...PNSQD.. KLMSEXYLGX QICNLVACEG
Tf4 DRFTESLHYY STLFDSLEGV ...PSGQD.. KVMSEVYLGK QICNVVACDG
18310 ERFTNALFHY ATMFDSLEAM HRCTSGRDIT DSLTEVYLRG EIFDIVCGEG
18652 SRFVETLDYY TAMFESIDAA R..PRDDKQR ISAEQHCVAR DIVNMIACEE
4818 PRFVETMNHY LAVFESIDVK L..ARDHKER INVEQHCLAR EVENLIACEG
21729 GRVSESCACY GALLESVEST V..PSTNSDR AKVE.EGIGR KLVNAVACEG
1110 PRFIEAYEYY SAVFESLDMT L..PRESQER MNVERQCLAR DIVNIVACEG
174 ..... .RXFDSLEHD A..SKGEPRE DERGRXCLAR NIVNIVXCKX
601
650

FIG. 15Q

```
  Scr  PSRSGEVKF. .....ESWRE KMQQCGFKGI SLAG..NAAT QATLLLGMFP
 3989  .RHKA..Q.. ...KAIAWRS VFAATGFKPV QLSN..LAEA QADCLLKRVQ
12398  DNNKPGTRFG LMEEKEQWRV LMEKAGFEPV KPSN..YAVS QAKLLLWNYN
 4871  .RSRPIER.. ...PMMTWQA MFLQMGFSPV THSN..FTES QAECLVQRTP
11846  AARXERHE.. ...PLSRWRD RLTRAGLSAV PLG....SNA ..........
 2504  .......... .......... .......... .......... ..........
 3935  FERRERHE.. ...KLEKWSQ RIDLAGFGNV PLSY..YAML QARRLLQGCG
11261  AERFARPE.. ...TYKQWRV RILRAGFKPA TIS....KQI MKEAKEIVRK
  713  LERMERPE.. ...TYKQWQV RNQRVGFKQL PLN....QDM MKRARXEGQV
10964  WERVERPE.. ...TYKQWHV RAMRSGLVQV PFD....PSI MKTSLHKVHT
23196  TERVESRE.. ...TYKQWQA RLIRAGFRQL PLE....KEL MQNLKLKIEN
  Tf1  PDRVERHE.. ...TLSQWGN RFGSSGLAPA HLGS...NAF KQASMLLSVF
  Tf4  PDRVERHE.. ...TLSQWRN RFGSAGFAAA HIGS...NAF KQASMLLALF
18310  SARTERHE.. ...LFGHWRE RLTYAGLTQV WFDPDEVDTL KDQLIHVTSL
18652  SERVERHE.. ...VLGKWRV RMMMAGFTGW PVSTSAAFAA SE....MLK.
 4818  VEREERHE.. ...PLGKWRS RFHMAGFKPY PLSSYVNATI KG....LLE.
21729  IDRIERCE.. ...VFGKWRM RMSMAGFELM PLSEKIAESM KS....RGNR
 1110  EERIERYE.. ...AAGKWRA RMMMAGFNPK PMSAKVTNNI QN....LIKQ
  174  EERIERYE.. ...VTGKWRA RMMMAGFSPR PMSGRVTSNI ES....LIKR
       651                                                  700
```

FIG. 15R

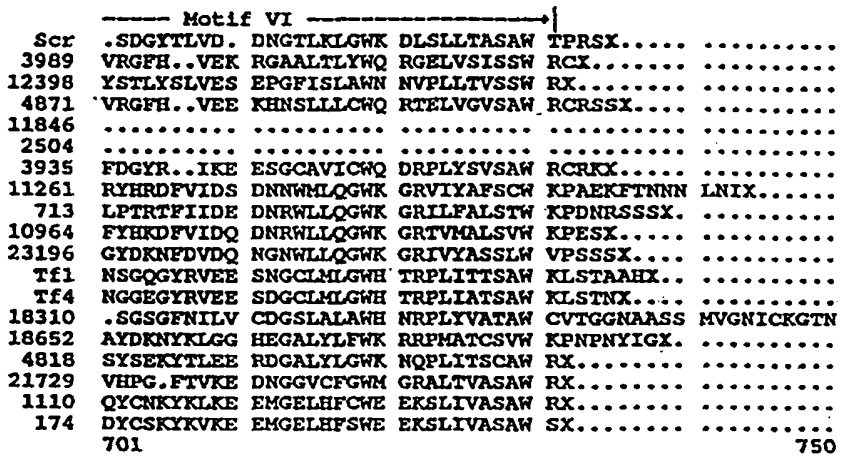

Motif VI
Scr .SDGYTLVD. DNGTLKLGWK DLSLLTASAW TPRSX..... .....
3989 VRGFH..VEK RGAALTLYWQ RGELVSISSW RCX..... .....
12398 YSTLYSLVES EPGFISLAWN NVPLLTVSSW RX..... .....
4871 VRGFH..VEE KHNSLLLCWQ RTELVGVSAW RCRSSX.... .....
11846 ....
2504 ....
3935 FDGYR..IKE ESGCAVICWQ DRPLYSVSAW RCRKX..... .....
11261 RYHRDFVIDS DNNWMLQGWK GRVIYAFSCW KPAEKFTNNN LNIX.....
713 LPTRTFIIDE DNRWLLQGWK GRILFALSTW KPDNRSSSX. .....
10964 FYHKDFVIDQ DNRWLLQGWK GRTVMALSVW KPESX..... .....
23196 GYDKNFDVDQ NGNWLLQGWK GRIVYASSLW VPSSSX.... .....
Tf1 NSGQGYRVEE SNGCLMLGWH TRPLITTSAW KLSTAAHX.. .....
Tf4 NGGEGYRVEE SDGCLMLGWH TRPLIATSAW KLSTNX.... .....
18310 .SGSGFNILV CDGSLALAWH NRPLYVATAW CVTGGNAASS MVGNICKGTN
18652 AYDKNYKLGG HEGALYLFWK RRPMATCSVW KPNPNYIGX. .....
4818 SYSEKYTLEE RDGALYLGWK NQPLITSCAW RX..... .....
21729 VHPG.FTVKE DNGGVCFGWM GRALTVASAW RX..... .....
1110 QYCNKYKLKE EMGELHFCWE EKSLIVASAW RX..... .....
174 DYCSKYKVKE EMGELHFSWE EKSLIVASAW SX..... .....
701
750

FIG. 15S

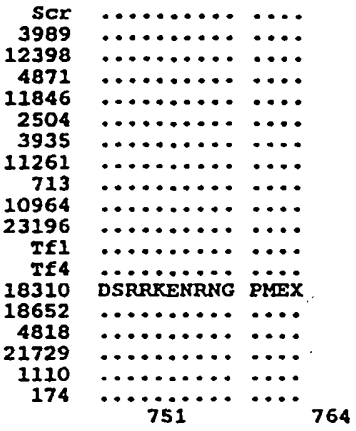

Scr ....
3989 ....
12398 ....
4871 ....
11846 ....
2504 ....
3935 ....
11261 ....
713 ....
10964 ....
23196 ....
Tf1 ....
Tf4 ....
18310 DSRRKENRNG PMEX
18652 ....
4818 ....
21729 ....
1110 ....
174 ....
751
764

FIG. 16A

SRPa1 (1110)

CTTTGTCAATGGTAAATGAGCTGAGGCAGATAGTTTCTATCCAAGGAGACCCTTCTCAGA
GAATCGCAGCTTACATGGTGGAAGGTCTAGCTGCAAGAATGGCCGCTTCAGGAAAATTCA
TCTACAGAGCATTGAAATGCAAAGAGCCTCCTTCGGATGAGAGGCTTGCAGCTATGCAAG
TCCTGTTTGAAGTCTGCCCTTGTTTCAAGTTCGGGTTTTTAGCAGCTAATGGTGCGATAC
TTGAAGCAATCAAAGGTGAAGAAGAAGTTCACATAATCGATTTCGATATAAACCAAGGGA
ACCAATACATGACACTGATACGAAGCATTGCTGAGTTGCCTGGTAAACGACCTCGCCTGA
GGTTAACAGGAATTGATGACCCTGAATCAGTCCAACGCTCCATTGGAGGGCTAAGAATCA
TCGGTCTAAGACTCGAGCAACTCGCAGAGGATAATGGAGTATCCTTCAAATTCAAAGCAA
TGCCTTCAAAGACTTCGATTGTCTCTCCATCAACACTCGGTTGCAAACCAGGAGAAACCT
TAATAGTGAACTTTGCATTCCAACTTCACCACATGCCTGACGAGAGTGTCACAACAGTAA
ACCAGCGGGACGAGCTACTTCACATGGTCAAAAGCTTAAACCCAAAGCTTGTCACGGTCG
TTGAACAAGACGTGAACACAAACACTTCACCGTTCTTTCCCAGATTCATAGAGGCTTACG
AATACTACTCAGCAGTTTTCGAGTCTCTAGACATGACACTTCCAAGAGAAAGCCAAGAGA
GGATGAATGTAGAAAGACAGTGTCTCGCTAGAGACATAGTCAACATTGTTGCTTGCGAAG
GAGAAGAACGGATAGAGAGATACGAGGCTGCGGGAAAATGGAGAGCAAGGATGATGATGG
CTGGATTCAATCCAAAACCAATGAGTGCTAAAGTAACCAACAATATACAAAACCTGATAA
AGCAACAATATTGCAATAAGTACAAGCTTAAAGAAGAAATGGGTGAGCTCCATTTTTGCT
GGGAGGAGAAAAGCTTAATCGTTGCTTCAGCTTGGAGGTAAGATAAGTGACAAGAGCATA
TAGTCTTTATGTTTCATAAAACATAATTATGTTTTTACTGTAATCTTGGGTTATTGTGTA
ACTGGTTAAATCATCTCCATGTATTATTACCAGAGGTTAGGGGTGATCACAGGTACTAAA
AGCTAATCTAACACTTATGGAAGAATTTTTCTTTCTTTTTTTTCCCTATTATATAAAAAT
AATTAGAGTTTTGGTTCTAAACCTATTTGCTAAGTGTGAATGAGTCTTTACATGTTCATA
TTTCAGTTCAAATGGTTAAATTTGTTAAGGTTCTCACTTAAAAAAAAA

FIG. 16B

SRPa3 (3935)

GCTATGGAAGGAGAGAAGATGGTTCATGTGATTGATCTCGATGCTTCTGAGCCAGCTCAA
TGGCTTGCTTTGCTTCAAGCTTTTAACTCTAGGCCTGAAGGTCCACCTCATTTGAGAATC
ACTGGTGTTCATCACCAGAAGGAAGTGCTTGAACAAATGGCTCATAGACTCATTGAGGAA
GCAGAGAAACTCGATATCCCGTTTCAGTTTAATCCCGTTGTGAGTAGGTTAGACTGTTTA
AATGTAGAACAGTTGCGGGTTAAAACAGGAGAGGCCTTAGCCGTTAGCTCGGTTCTTCAA
TTGCATACCTTCTTGGCCTCTGATGATGATCTCATGAGAAAGAACTGCGCTTTACGGTTT
CAGAACAACCCTAGTGGAGTTGACTTGCAGAGAGTTCTAATGATGAGCCATGGCTCTGCA
GCTGAGGCACGTGAGAATGATATGAGTAACAACAATGGGTATAGCCCTAGCGGTGACTCG
GCCTCATCTTTGCCTTTACCAAGTTCAGGAAGGACTGATAGCTTCCTCAATGCTATTTGG
GGTTTGTCTCCAAAGGTCATGGTGGTCACTGAGCAAGACTCAGACCACAACGGCTCCACA
CTAATGGAGAGGCTATTAGAATCACTTTACACCTACGCAGCATTGTTTGATTGCTTGGAA
ACAAAAGTTCCAAGAACGTCTCAAGATAGGATCAAAGTGGAGAAGATGCTCTTCGGGGAG
GAGATCAAGAACATCATATCCTGCGAGGGATTTGAGAGAAGAGAAAGACACGAGAAGCTT
GAGAAATGGAGCCAGAGGATCGATTTGGCTGGTTTTGGGAATGTTCCTCTTAGCTATTAT
GCGATGTTGCAGGCTAGGAGATTGCTTCAAGGGTGCGGTTTTGATGGGTATAGAATCAAG
GAAGAGAGCGGGTGCGCAGTAATTTGCTGGCAAGATCGACCTCTATACTCGGTATCAGCT
TGGAGATGCAGGAAGTGAATGATATATTACAGTTTGTCTTCTATTTTGGTTATGAGCAGA
GTCCCTTTCTTTTTTGTATACATGGGGACACAATCTTAGTTGTTTTGTGATGGTGACTTT
CTGTCTCTTTATGCTATTTTGGCTTAAATGCTTCTACTGCCTCTGCATGTAAAGCCTTTG
TGTGTTGGTTCAATTTGGTCTGGTGTGGGTGTAATACCAAACCAAATCCAATTTGAGCTG
AAGATAACTAATTTGATGATCGGCTCGTGCC

FIG. 16C

SRPa4 (4818)

GGCACGAGCCCAACGGGTCCTGAGCTTCTTACTTATATGCATATCTTGTATGAAGCCTGC
CCTTATTTCAAATTCGGTTATGAATCTGCTAATGGAGCTATAGCTGAAGCTGTGAAGAAC
GAAAGTTTTGTGCACATTATCGATTTCCAGATTTCTCAAGGTGGTCAATGGGTGAGTTTG
ATCCGTGCTCTTGGTGCTAGACCTGGTGGACCTCCGAACGTTAGGATAACGGGAATTGAT
GATCCGAGATCATCGTTTGCTCGTCAAGGAGGACTTGAGTTAGTTGGACAAAGACTTGGG
AAGCTAGCTGAAATGTGCGGTGTTCCGTTTGAGTTCCATGGAGCTGCTTTATGCTGCACG
GAAGTCGAAATCGAGAAGCTAGGAGTTAGAAATGGAGAAGCGCTCGCGGTTAACTTCCCG
CTTGTTCTTCACCACATGCCTGATGAGAGTGTAACTGTGGAGAATCACAGAGATAGATTG
TTGAGATTGGTCAAACACTTGTCACCAAACGTTGTGACTCTGGTTGAGCAAGAAGCGAAT
ACAAACACTGCGCCGTTTCTTCCCCGGTTTGTCGAGACAATGAACCATTACTTGGCAGTT
TTCGAATCAATAGATGTGAAACTCGCTAGAGATCACAAGGAAAGGATCAATGTTGAGCAG
CATTGTTTGGCTAGAGAGGTTGTGAATCTTATAGCTTGTGAAGGTGTTGAAAGAGAAGAG
AGGCACGAGCCACTAGGGAAATGGAGGTCTCGGTTTCACATGGCGGGATTTAAACCGTAT
CCTTTGAGCTCGTATGTGAACGCAACAATCAAAGGATTGCTTGAGAGTTATTCAGAGAAG
TATACACTTGAAGAAAGAGATGGAGCATTGTATTTAGGATGGAAGAATCAACCTCTTATC
ACTTCTTGTGCTTGGAGGTAACTAATAAAAACCTTGTTCGGTTTCAGAAGAGATTAGAAA
CTTCTTTTAAAGTTTGCAGAATCTGTTTGTAAAAGTAAAACTCATGCATGATCCGNAGGA
ACAAGTTGTCAAATGTTGTAGTAGTAAGTGATATGTTGATGACCCAAAAAAAAAAAAAAA
AAAAA

FIG. 16D

SRPa5 (4871)

GCGGCTATCTTCTACGGCCACCACCACCATACACCTCCGCCGGCAAAGCGGCTCAACCCT
GGTCCCGTGGGGATAACAGAGCAGCTGGTTAAGGCAGCAGAGGTCATAGAGAGCGACACG
TGTCTAGCTCAGGGGATATTGGCGCGGCTCAATCAACAGCTCTCTTCTCCCGTCGGGAAG
CCATTAGAAAGAGCAGCTTTTTACTTCAAAGAAGCTCTCAATAATCTCCTTCACAACGTC
TCCCAAACCCTAAACCCTTATTCCCTCATCTTCAAGATCGCTGCTTACAAATCCTTCTCA
GAGATCTCTCCCGTTCTTCAGTTCGCCAACTTTACCTCCAACCAAGCCCTCTTAGAGTCC
TTCCATGGCTTCCACCGTCTCCACATCATCGACTTCGATATCGGCTACGGTGGCCAATGG
GCTTCCCTCATGCAAGAGCTTGTTCTCCGCGACAACGCCGCTCCTCTCTCCCTCAAGATC
ACCGTTTTCGCTTCTCCGGCGAACCACGACCAGCTCGAACTTGGCTTCACTCAAGACAAC
CTCAAGCACTTCGCCTCTGAGATCAACATCTCCCTTGACATCCAAGTTTTGAGCTTAGAC
CTCCTCGGCTCCATCTCGTGGCCTAACTCGTCGGAGAAAGAAGCTGTCGCCGTTAACATC
TCCGCCGCGTCCTTCTCGCACCTCCCTTTGGTCCTCCGTTTCGTGAAGCATCTATCTCCG
ACGATCATCGTCTGCTCCGACAGAGGATGCGAGAGGACGGATCTGCCCTTCTCTCAACAG
CTCGCCCACTCGCTGCACTCACACACCGCTCTCTTCGAATCCCTCGACGCCGTCAACGCC
AACCTCGACGCAATGCAGAAGATCGAGAGGTTTCTTATACAGCCGGAGATAGAGAAGCTG
GTGTTGGATCGTAGCCGTCCGATAGAAAGGCCGATGATGACGTGGCAAGCGATGTTTCTA
CAGATGGGTTTCTCACCGGTGACGCACAGTAACTTCACGGAGTCTCAAGCCGAGTGTTTA
GTCCAACGGACGCCAGTGAGAGGCTTTCACGTCGAGAAGAAACATAACTCACTTCTCCTA
TGTTGGCAAAGGACAGAACTCGTCGGAGTTTCAGCATGGAGATGTCGCTCCTCCTGATTT
CCACCGGAGTTTCAATTATTAAAAAAATATTTTCCTTAATTCAATTTATCTTAAATGACA
AATTTTTAGTTTCTGATTTTATTTTGCTCAGTGCGATGGATTTTTAAATTTAAGTTTCAC
ACAAATATATAAATTTTTG

FIG. 16E

SRPa6 (12398)

AATCGCTTGAACCGAATTTGGATCGAGATTCGAAAGAAAGGCTGAGAGTGGAGAGAGTGC
TGTTCGGTAGGAGGATTATGGATTTGGTCCGATCAGATGATGATAATAATAAACCGGGAA
CCCGGTTTGGGTTAATGGAGGAGAAAGAACAATGGAGAGTGTTGATGGAGAAAGCTGGAT
TTGAGCCGGTTAAACCGAGTAATTACGCGGTTAGCCAAGCGAAGCTGCTACTATGGAACT
ACAATTATAGTACATTGTATTCACTTGTTGAATCGGAGCCAGGTTTCATCTCCTTGGCTT
GGAACAATGTGCCTCTCCTCACCGTTTCCTCTTGGCGTTGACTACTTGGTCCGATAAGTT
AATCTAGTATTTTGAGTTAGCTTTTAGAATTGAATTGTTTGGGGTTAGATTTGGATGTTT
AATTAGTCTCTAGCCTATTCTCTTACTCTTTTTTGTCTAGTGCTTGGAGTGATGATGGTT
TGTCGTTTATGTTCATTTGTAATATATATTGTATGTAACATTTGACTAAAAAAAAAAAAA
AAAAAAAA

FIG. 16F

SRPa7 (21729/3635/17410)

AAAGACTTTAGCAGATTTTCAAGCGGCTCAGAACATCAACAACAACAACAACAACAACCG
TTTTATAGTCAAGCAGCTCTCAACGCTTTTCTTTCAAGGTCTGTGAAGCCTCGAAATTAT
CAGAATTTTCAATCTCCGTCGGCCGATGATTGATCTCACGTCGGTGAATGATATGAGTTT
GTTTGGTGGTTCTGGTTCATCTCAGCGTTACGGTTTACCGGTTCCCAGGTCTCAGACGCA
ACAGCAACAATCGGATTACGGTTTATTTGGTGGGATCCGAATGGGAATCGGGTCGGGTAT
TAATAATTATCCAACATTAACCGGCGTTCCGTGTATTGAACCGGTTCAAAACCGGGTTCA
TGAATCGGAGAACATGTTGAATAGTTTAAGAGAGCTTGAGAAACAGCTTTTAGATGATGA
CGATGAGAGTGGTGGTGATGATGACGTGTCAGTTATAACAAATTCAAATTCCGATTGGAT
TCAAAATCTCGTGACTCCGAACCCGAACCCGAACCCGGTTTTGTCTTTTTCACCGAGCTC
TTCTTCTTCGTCTTCTTCGCCTTCTACAGCTTCGACGACGACATCGGTATGTTCTAGGCA
AACGGTTATGGAAATCGCGACGGCGATCGCGGAAGGGAAAACAGAGATAGCGACGGAGAT
TTTGGCGCGTGTTTCTCAAACGCCTAATCTTGAGAGGAATTCAGAGGAGAAGCTTGTTGA
TTTCATGGTGGCTGCGCTTCGATCGAGGATAGCTTCTCCAGTGACGGAATTGTATGGGAA
GGAGCATTTAATCTCGACTCAATTGCTCTACGAGCTCTCTCCTTGTTTCAAACTCGGTTT
CGAGGCCGCGAATCTCGCCATTCTCGACGCCGCCGATAACAACGACGGTGGAATGATGAT
ACCGCACGTTATCGATTTCGATATCGGAGAAGGTGGACAATACGTTAACCTTCTCCGTAC
ATTATCCACGCGCCGGAATGGTAAAAGTCAGAGTCAGAATTCTCCGGTGGTTAAGATCAC
CGCCGTGGCGAACAACGTTTACGGATGTTTAGTCGATGACGGTGGAGAAGAGAGGTTAAA
AGCCGTCGGAGATTTGTTGAGCCAACTCGGTGATCGACTCGGTATCTCCGTAAGTTTCAA
CGTGGTGACGAGTTTACGACTCGGTGATCTGAATCGTGAATCTCTCGGGTGTGATCCCGA
CGAGACTTTGGCTGTGAACTTAGCTTTCAAGCTTTATCGTGTTCCCGACGAAAGCGTATG
CACGGAGAATCCAAGAGACGAACTTCTCCGGCGCGTGAAGGGACTTAAACCGCGCGTGGT
TACTCTAGTGGAGCAAGAAATGAATTCGAATACGGCGCCGTTTTTAGGGAGAGTGAGTGA
GTCATGCGCGTGTTACGGTGCGTTGCTTGAGTCGGTCGAGTCTACGGTTCCTAGTACGAA
TTCCGACCGTGCCAAAGTTGAGGAAGGAATTGGCCGGAAGCTAGTAAACGCGGTGGCGTG
CGAAGGAATCGATCGTATAGAGCGGTGCGAGGTGTTCGGGAAATGGCGAATGCGGATGAG
CATGGCTGGGTTTGAGTTAATGCCATTGAGTGAGAAGATAGCGGAGTCGATGAAGAGTCG
TGGAAACCGAGTCCACCCGGGCTTTACCGTTAAAGAAGATAACGGAGGTGTGTGCTTTGG
TTGGATGGGACGGGCACTCACTGTCGCATCCGCTTGGCGTTAACTTCACACACTCTTTTT
TTTCTTCTTATTATTACCATATTATTATTAATTTTCGAGATTATTCTGATATTATTATCA
TTGTGATTTTCCGTTTCGAAAAGTGTAGGAATCTTATGTAACAAAGAAAAAAAAAAGACT
TTTATGTTTTTCTAATAATAAAAGAAAGAGTGATTGGGTTCAAAAAAAAAAAAAAAAAAA
AAAAAAAA

FIG. 16G

SRPa8 (10964)

TGCATACAACGCACCGTTTTTCGTAACACGGTTTCGCGAAGCTCTATTTCATTTCTCCTC
GATTTTTGACATGCTTGAGACAATTGTGCCACGAGAAGACGAAGAGAGGATGTTCCTTGA
GATGGAGGTCTTTGGGAGAGAGGCACTGAATGTGATTGCTTGCGAAGGTTGGGAAAGAGT
GGAGAGGCCTGAGACATACAAGCAGTGGCACGTACGGGCTATGAGGTCAGGGTTGGTGCA
GGTTCCATTTGACCCAAGCATTATGAAGACATCGCTGCATAAGGTCCACACATTCTACCA
CAAGGATTTTGTGATCGATCAAGATAACCGGTGGCTCTTGCAAGGCTGGAAGGGAAGAAC
TGTCATGGCTCTTTCTGTTTGGAAACCAGAGTCCAAGGCTTGACCGAGAAATCCTCGTTG
GCATATGAGAGACCATCTCTTGATTTTCTTCCTGTGTAATTCCCAGAGACAGAATTACAG
ATGTAAGAAGAGAATGCTGCACAAAGAACTTGTTCAAAGATAATATTGATGTAAGTCCTG
TTTTATAACTTTCTAGCTGTGTTTTTGTTGTTTCTCAGCTAGATTCTCCTAACGGTATTC
TTGTAGCTAGGGTGATCAGATTGTTTGTATATTGCTAGCAGAGTTAGTTTGTCTAGATTG
TAACACATATAAGAGGAAGCTTAGAGTTTCTATGGTTTAAAGAGAAGTTTTTTCCTTCTC
CAATGTAAAAAAAAAAAAAAAAAA

FIG. 16H

SRPa10 (11261)

AAAAAATGGGAAACCATCACTCTTGATGAACTTATGATCAATCCAGGAGAGACAACGGTC
GTCAACTGCATTCATCGGTTACAATACACTCCTGATGAAACTGTGTCATTAGACTCTCCA
AGAGACACGGTTCTGAAGCTATTCAGAGATATCAATCCTGACCTCTTTGTGTTTGCAGAG
ATTAACGGAATGTACAACTCTCCTTTCTTCATGACGAGGTTCCGAGAAGCGCTTTTTCAT
TACTCTTCACTCTTTGACATGTTTGACACCACAATACACGCAGAGGATGAGTACAAAAAC
AGGTCACTGTTGGAGAGAGAGTTACTTGTGAGAGACGCGATGAGCGTGATTTCCTGCGAG
GGTGCAGAGCGGTTTGCGAGGCCTGAAACCTACAAGCAATGGCGAGTTAGGATTTTGAGA
GCCGGGTTTAAGCCAGCAACTATTAGCAAACAGATCATGAAGGAGGCTAAGGAAATTGTG
AGGAAACGTTACCATAGAGATTTTGTGATCGATAGCGATAACAATTGGATGCTTCAAGGA
TGGAAAGGAAGAGTCATCTATGCTTTTTCTTGCTGGAAACCTGCTGAGAAGTTCACAAAC
AATAATTTAAACATCTGAAAAATGTTACTTCTCAATTACATCATTTTTGTTTCCCAATGG
TTTTGTAGAATATGTTTGATCCCGTGAGTGGATGCAACTCTTTTTTCCTGCAAGTACATA
TTGTATTCAAATCCTTGTGGAAATGATAAATTGTTTAATCAAAAAAAAAAAAAAA

FIG. 16I

SRPa11 (18652)

GCGAATGTTGAGATCTTGGAAGCAATAGCTGGGGAAACCAGAGTCCACATTATCGATTTT
CAGATTGCACAGGGATCACAATACATGTTTTTGATTCAGGAGCTTGCGAAACGCCCTGGT
GGGCCGCCGTTGCTGCGTGTGACGGGTGTGGATGATTCACAGTCCACCTATGCTCGTGGG
GGAGGACTCAGCTTGGTAGGTGAGAGGCTTGCAACTTTGGCGCAGTCATGTGGTGTCCCG
TTTGAGTTTCACGATGCCATCATGTCTGGGTGCAAGGTGCAGCGGGAACATCTCGGGTTG
GAACCTGGCTTTGCTGTTGTTGTGAACTTCCCATATGTATTACACCACATGCCAGACGAG
AGCGTAAGTGTTGAAAAATACAGAGACAGGCTGCTGCATCTGATCAAGAGCCTCTCCCCA
AAACTGGTTACTCTAGTAGAGCAAGAATCCAACACAAACACCTCGCCATTGGTGTCACGG
TTTGTGGAAACACTGGATTACTACACAGCGATGTTTGAGTCGATAGATGCAGCACGGCCA
CGGGATGATAAGCAGAGAATCAGCGCAGAACAACACTGTGTAGCAAGAGACATAGTGAAC
ATGATAGCATGTGAGGAGTCAGAGAGAGTAGAGAGACACGAGGTACTGGGGAAATGGAGG
GTCAGAATGATGATGGCTGGGTTCACGGGTTGGCCGGTCAGCACATCTGCAGCGTTTGCA
GCGAGTGAGATGCTGAAAGCTTATGACAAAAACTACAAACTGGGAGGCCATGAAGGAGCG
CTCTACCTCTTCTGGAAGAGACGACCCATGGCTACATGTTCCGTGTGGAAGCCAAACCCA
AACTATATTGGGTAAGTTATAGTGATGATGGTTACTTGAGTGGATAAAGAAGAGCACAAC
AAAAACACATCTGTCGCTGTAAATTTTTTAGGATGTGCAATGATGTTTTAAGTTGTAACA
CAACCTAAGTTATATATGTATACAAACCAAACCTGGTGGTTGTTTTTCTCTTGTAAATTG
TCATGTGGTTGTGGGTGGGAAGCTAGTAATGAAATATAACCAAAACATTGATTAGGTCAA
AAAAAAAAAAAAA

FIG. 16J

SRPa12 (23196)*

TCTTACTCAAGGTTCTTCTTTGTCATCTTGTTGCCGAATCCACAAAGAGGAGAATAAAGA
TTCGACCTTTATTAGATATTAACGACTCTGGATTTTTGGGTTTTTGGAGTTGGATCCACA
TGGGTTCTTATCCGGATGGATTCCCTGGATCCATGGACGAGTTGGATTTCAATAAGGACT
TTGATTTGCCTCCCTCCTCAAACCAAACCTTAGGTTTAGCTAATGGGTTCTATTTAGATG
ACTTAGATTTCTCATCCTTGGATCCTCCAGAGGCATATCCCTCCCAGAACAACAACAACA
ACAACATCAACAACAAAGCTGTAGCAGGAGATCTGTTATCATCTTCATCTGATGACGCTG
ATTTCTCTGATTCTGTTTTGAAGTATATAAGCCAAGTTCTTATGGAAGAGGATATGGAAG
AGAAGCCTTGTATGTTTCATGATGCTTTGGCTCTTCAAGCTGCTGAGAAATCTCTCTATG
AGGCTCTTGGTGAGAAAGACCCTTCTTCGTCTTCTGCTTCTTCTGTGGATCATCCTGAGA
GATTGGCTAGTCATAGCCCTGACGGTTCTTGTTCAGGTGGTGCTTTTAGTGATTACGCTA
GCACCACTACCACTACTTCCTCTGATTCTCACTGGAGTGTTGATGGTTTGGAGAATAGAC
CTTCTTGGTTACATACACCTATGCCGAGTAATTTTGTTTTCCAGTCTACTTCTAGGTCCA
ACAGTGTCACCGGTGGTGGTGGTGGTGGTAATAGTGCGGTTTACGGTTCAGGTTTTGGCG
ATGATTTGGTTTCGAATATGTTTAAAGATGATGAATTGGCTATGCAGTTCAAGAAAGGGG
TTGAGGAAGCTAGTAAGTTCCTTCCTAAGTCTTCTCAGCTCTTTATTGATGTGGATAGTT
ACATCCCTATGAATTCTGGTTCCAAGGAAAATGGTTCTGAGGTTTTTGTTAAGACGGAGA
AGAAAGATGAGACAGAGCATCATCATCATCATAGCTATGCACCACCACCCAACAGATTAA
CTGGTAAGAAAAGCCATTGGCGCGACGAAGATGAAGATTTCGTTGAAGAAAGAAGTAACA
AGCAATCAGCTGTTTATGTTGAGGAAAGCGAGCTTTCTGAAATGTTTGATAACATGTTCC
TATGTGGCCCTGGGAAACCTGTATGCATTCTTAACCAGAACTTTCCTACAGAATCCGCTA
AAGTCGTGACCGCACAGTCAAATGGAGCAAAGATTCGTGGGAAGAAATCAACTTCTACTA
GTCATAGTAACGATTCTAAGAAAGAAACTGCTGATTTGAGGACTCTTTTGGTGTTATGTG
CACAAGCTGTATCAGTGGATGATCGTAGAACCGCCAACGTTTAGCTAAGGCAGATACGAG
AGCATTCTTCGCCTCTAGGCAATGGTTCAGAGCGGTTGGCTCATTATTTTGCAAATAGTC
TTGAAGCACGCTTAGCTGGGACCGGTACACAGATCTACACCGCTTTATCTTCGAAGAAAA
CGTCTGCAGCAGACATGTTGAAGGCTTACCAGACATACATGTCGGTCTGCCCTTTCAAGA
AAGCTGCTATCATATTTGCTAACCACAGCATGATGCGTTTCACTGCAAACGCCAACACGA
TCCACATAATAGATTTCGGAATATCTTACGGTTTTCAGTGGCCTGCTCTGATTCATCGCC
TCTCGCTCAGCAGACCTGGTGGTTCGCCTAAGCTTCGAATTACCGGTNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNGAGTTCAGGAGACAGGTCATCGCTTGGCTCGATACT
GTCAGCGACACAATGTTCCGTTTGAGTACAACGCAATTGCTCAGAAATGGGGAAACGATC
CAAGTCGAAGACTTAAAGCTTCGACAAGGAGAGTATGTGGTTGTGAACTCTTTGTTCCGT
TTCAGGAACCTTCTAGATGAGACCGTTCTGGTAAACAGCCCGAGAGATGCAGTTTTGAAG
CTGATAAGAAAAATAAACCCGAATGTCTTCATTCCAGCGATCTTAAGCGGGAATTACAAC
GCGCCATTCTTTGTCACGAGGTTCAGAGAAGCGTTGTTTCATTACTCGGCTGTGTTTGAT
ATGTGTGACTCGAAGCTAGCTAGGGAAGACGAGATGAGGCTGATGTATGTGTTTGAGTTT
TATGGGAGAGAGATTGTGAATGTTGTGGCTTCTGAAGGAACAGAGAGAGTGGAGAGCCGA
GAGACATATAAGCAGTGGCAGGCGAGACTGATCCGAGCCGGATTTAGACAGCTTCCGCTT
GAGAAGGAACTGATGCAGAATCTGAAGTTGAAAATCGAAAACGGGTACGATAAAAACTTC
GATGTTGATCAAAACGGTAACTGGTTACTTCAAGGGTGGAAAGGTAGAATCGTGTATGCT
TCATCTCTATGGGTTCCTTCGTCTTCATAGATGTTGTTTCTTACGTTCTAAGCGACTGGG
ATTTATGTAGGGCTTTTCTGTTGATAGTCTCTCGCCAACACGAGTGGATTAAGTTCAGAG
TTAGGGTTCTTGAACACTAGAATGTTGTTATATTATGCTTGTGACATAGCGTGTGTAAGA
GTGTAGCCTAAGAGATATAGTACTCATTGCATGATCTTTTGCTATATGTTNCATGT

FIG. 16K

SRPd1

TCTGCAGACAATTTTNAGGAGGCCAATACCATGCTATTGGAAATTTCAGAACTG
TCCACACCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTACTTCTCAGAGGN
AATGTCGGNNAGATTAGTTAGCTCCTGCTTAGGAATCTATGCTTCTCTTCCNGC
AACAGTGGTGCCTCCTCATGGTCAGAAAGTGGCCTCA

FIG. 16L

SRPq1

TCAACTGAGAATCTAGAAGATGCCAACAAGATGCTTCTGGAGATTTCTCAGTTA
TCAACACCGTTCNNCACTTCAGCACAGCGTGTGGCAGCATATTTCTCAGAAGCC
ATATCAGCAAGGTTGGTGAGTTCATGTCTAGGGATATACGCAACTTTGCCACAC
ACACACCAAAGCCACAAGGTAGCTTCAGCTTTTCAAGTGTTCAATGGTATTAGT
CCTTTAGTGGAGTTCTCACACTTCACAGCAAACCAAGCAATTCAAGAAGCCTTC
GAAAGAGAAGAGAGGGTGCACATCATAGATCTTGATATAATGCAAGGGTTG

FIG. 16M

SRPp1

TCTGCAGACAACTTTGAAGAAGCCAATACAATACTGCCTCAGATCACAGAACTC
TCCACCCCCTATNGCAACTCGGTGCAACGAGTGGCTGCCTATNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNTGCATAGGAATGTATTCTCCTCTCCCTCCT
ATTCACATGTCCCAGAGCCAGAAAATTGTGAAT

FIG. 17A

Partial DNA sequence of *ZCARECROW* gene

GATATCAGCATCATCAATTTTAAATGTAAGTTGGCAAAAGATCATGAGGGTTCTCATAGT
AATTTGGCCACAAGGTATGACACTGTCTCAATTGAGCAATCTAGTAGAGAAACTGATCCA
TCATATATTGCTCATATTGAAAGTGAAAAAGATATGCTCAAGAACCTAGTAGAGAAGCTA
AAAATTGAAAAATCTAGCTCTACTAGAAAAATATGATAGGTTGCCTGTTTCTCATGAAAA
TTTATTAGATAATCATATCATGGCTAGATGTCGCTCATGAGGTTGTTCTTGCTAGTTTAG
ATTCCTGTGGGCATTCATCTCTTTTAGATGCACTAACATGATAGGAAGTTTCTAATCTGG
TGCTTCACAATTCTGGTGATTCATGCTTCCTTCATTGCAATTGATATTGATGCTTGATTC
ATGCTTCAGTCACTTTGTGCGTTTAATTGGTATTGTATGTATCACTAGATTGTAGGGTGT
CTGCAACTAGTGTTTCACCATGTGGTTTTTTAGTATCATTCGTATTAGTTTCTAACTTTC
TATTGATATATTAAAGTGATAACTAGTTTTAGAAATATTCTCTTGTGCCATTAATGCTAC
AACTTGTTTTTAGCGTGTACGTTAGCATTATAATATTTCCTTATTATGAAAGCGGAAGAG
AAACGCGCCCAACCAGAGCATCCACGTCGTCTCATTTCACCTTCATCGTTGGATCATAGA
TGAGCGGTCCACGGTGAACTCCGTTTGCCTGCAAAACCACGTCCTCTACGCGCTGTTAAG
TAGCTTCTAGAAACATCACGATGTGTCCCGTCCATTCCTTTAGGAGGAGCCGGATCCGGC
GCCGCAGTCGCCCAAGGTCCCGACCGCCGCGGCCTCGGCCGCCGCCGCCAAGGAGCGGAA
GGAGGTGCAGCGGCGGAAGCAGCGCGACGAGGAGGGCCTCCACCTGCTGAGTGCTGACGC
TGCTGCTGCAGTGCGCGGAGGCCGTGAACGCGGACAACCTCGACGACGCGCACCAGACGC
TGCTGGAGATCGCGGAGCTGGCCACGCCGTTCGGCACCTCGACCCAGCGCGTGGCCGCCT
ACTTCGCGGAGGCCATGTCGGCGCGCGTCGTCAGCTCCTGCCTAGGCCTGTACGCGCCGC
TGCCGCCGGGCTCCCCCGCCGCGGCGCGCCTCCACGGCCGCGTGGCCGCCGCGTTCCAGG
TGTTCAACGGCATCAGCCCCTTCGTCAAGTTCTCGCACTTCACCGCCAACCAGGCCATCC
AGGAGGCGTTCGAGCGGGAGGAGCGTGTGCACATCATCGACCTCGACATCATGCAGGGGC
TGCAGTGGCCGGGCCTCTTCCACATCCTTGTCTCCCGCCCCGGCGGCCCGCCCAGGGTCA
GGCTCACCGGCCTGGGGGCGTCCATGGACGCGCTCGAGGCGACGGGGAAGCGCCTCTCCG
ACTTCGCCGACACGCTCGGCCTGCCCTTCGAGTTCTGCGCCGTCGCCGAGAAGGCCGGCA
ACGTTGACCCGCAGAAGCTGGGCGTCACGCGGCGGGAGGCCGTCGCCGTCCACTGGCCGC
ACCACTCGCTTTACGACGTCATCGGCTCCGACTCCAACACGCTCTGGCTCATCCAAAGGT
CCTCCATTTTCCTTCTCTGCCTTTCTTCCATGTCAAATCTTGATGCAATCATGACCACTT
TTCAGCTGCTGACATTGGATAATGTGAGCTTTACGGCAAGCATCAAGTCGTGGTAGTACA
TCCATTACAGCTATTTCTAAAATATTCTTCGGAGGTTTCCTGCTCATAGTAAAAAAAAAT
CGCGTTTTGAAGCTCAAAAGGCGATTTCTTCCGAGGTTTGCTGTTGAGCGCTATTTTGGA
AACCCCATTTTCTCAATTGATTTTTATTTTTTAAAGAAAAATTAGTTCATTTTTCTCTTG
TGAAATGGAGTCCCAAACTAACCCTAATATTAAAAAAAACGCGCTTTGGAGCTCAAAACG
CTCGTTGTTATGACCAACCAGCTTTATAGGTTTAAAAAGGTTGAATCTTGACAATGCTTT
TGAAAAGGTTGAATCTTGACAATGCTTTTGAGATGATACTGTAGTGTAGTCTGTAGTGGA
GCATCCTCCATGGTCTTTGGTGATCGAGAATTCCTGCAGCCCGGGGGATCC

FIG. 17B

Partial amino acid sequence of ZCARECROW protein

YQHHQFXMXVGKRSXGFSXXFGHKVXHCLNXAIXXRNXSIIYCSYXKXKRYAQEPSREAK
NXKIXLYXKNMIGCLFLMKIYXIIISWLDVAHEVVLASLDSCGHSSLLDALTXXEVSNLV
LHNSGDSCFLHCNXYXCLIHASVTLCVXLVLYVSLDCRVSATSVSPCGFLVSFVLVSNFL
LIYXSDNXFXKYSLVPLMLQLVFSVYVSIIIFPYYESGRETRPTRASTSSHFTFIVGSXM
SGPRXTPFACKTTSSTRCXVASRNITMCPVHSFRRSRIRRRSRPRSRPPRPRPPPPRSGR
RCSGGSSATRRASTCXVLTLLLQCAEAVNADNLDDAHQTLLEIAELATPFGTSTQRVAAY
FAEAMSARVVSSCLGLYAPLPPGSPAAARLHGRVAAAFQVFNGISPFVKFSHFTANQAIQ
EAFEREERVHIIDLDIMQGLQWPGLFHILVSRPGGPPRVRLTGLGASMDALEATGKRLSD
FADTLGLPFEFCAVAEKAGNVDPQKLGVTRREAVAVHWPHHSLYDVIGSDSNTLWLIQRS
SIFLLCLSSMSNLDAIMTTFQLLTLDNVSFTASIKSWXYIHYSYFXNILRRFPAHSKKKS
RFEAQKAISSEVCCXALFWKPHFLNXFLFFKEKLVHFSLVKWSPKLTLILKKTRFGAQNA
RCYDQPALXVXKGXILTMLLKRLNLDNAFEMILXCSLXWSILHGLWXSRIPAARGI

FIG. 18

```
            302                                                  349
SCR         SADNLEEANKLLLEISQLSTPYGTSAQRVAAYFSEAMSARLLNSCLGI
SRPd1       SADNFxEANTMLLEISELSTPxxxxxxxxxxYFSxxMSxRLVSSxLxI
SRPg1       STENLEDANKMLLEISQLSTGxxxxxxxxxxxxxxxxxxxxxxxxxSCLGI
SRPp1       SADNFEEANTILPQITELSTPYxNSVGRVAAYxxxxxxxxxxxxCIGM

            350                                                  396
SCR         YAALPSRWMPQTH-SLKMVSAFQVFNGISPLVKFSHFTANQAIQEAFE
SRPd1       YASLPATVVP--PHGQKVAS
SRPg1       YATLP-----HTHQSHKVASAFQVFNGISPLVEFSHFTANQAIQEAFE
SRPp1       YSPLPPIxMSQ---SQKIVN

            397          412
SCR         KEDSVHIIDLDIMQGL
SRPg1       REERVHIIDLDIMQGL
```

FIG. 19

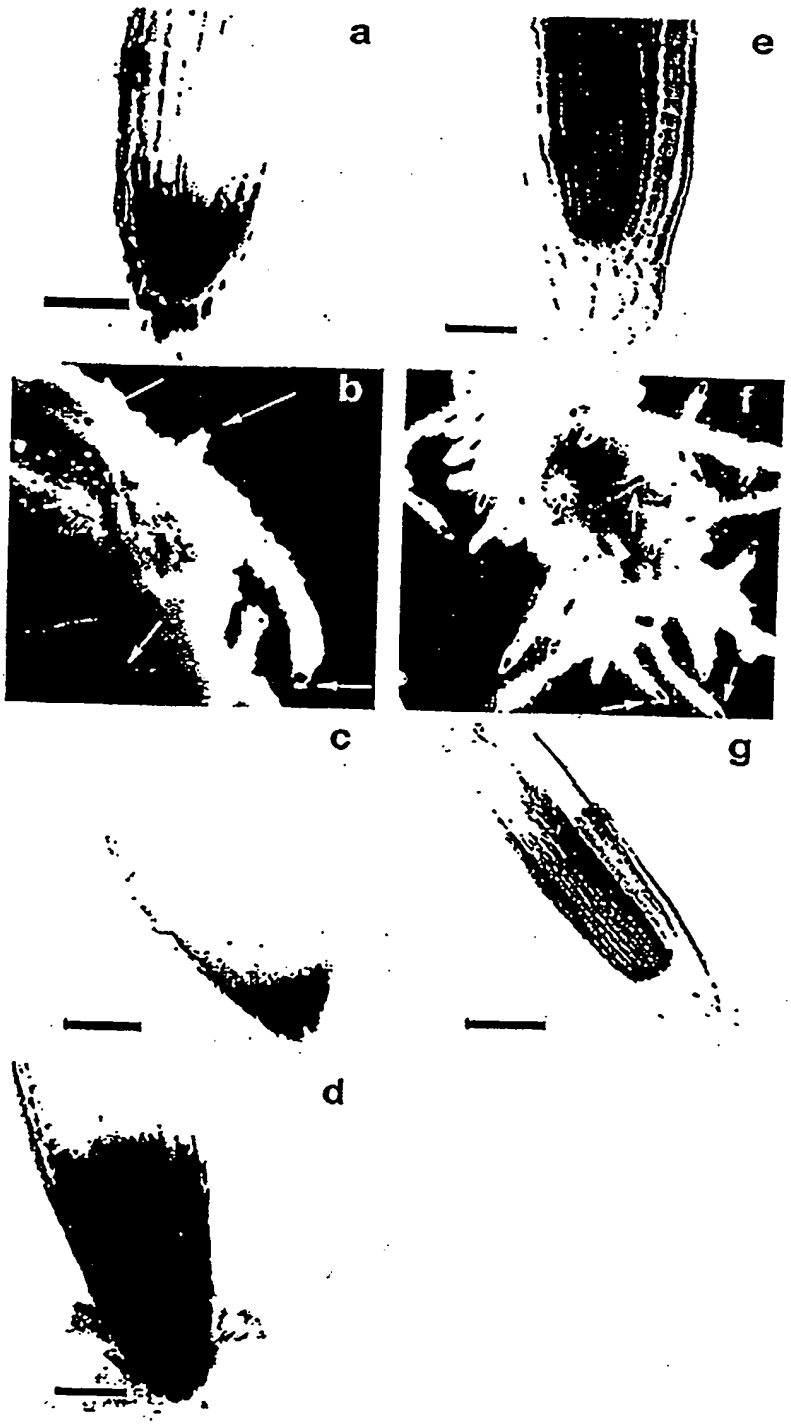


FIG. 20

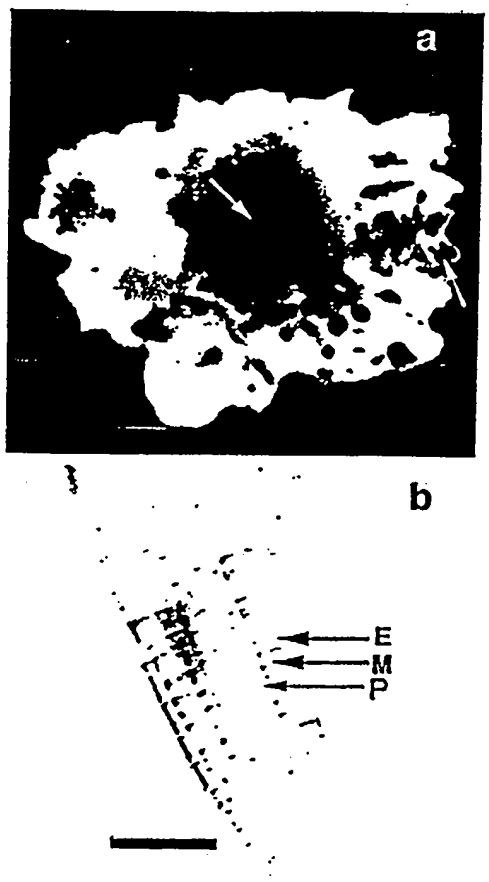
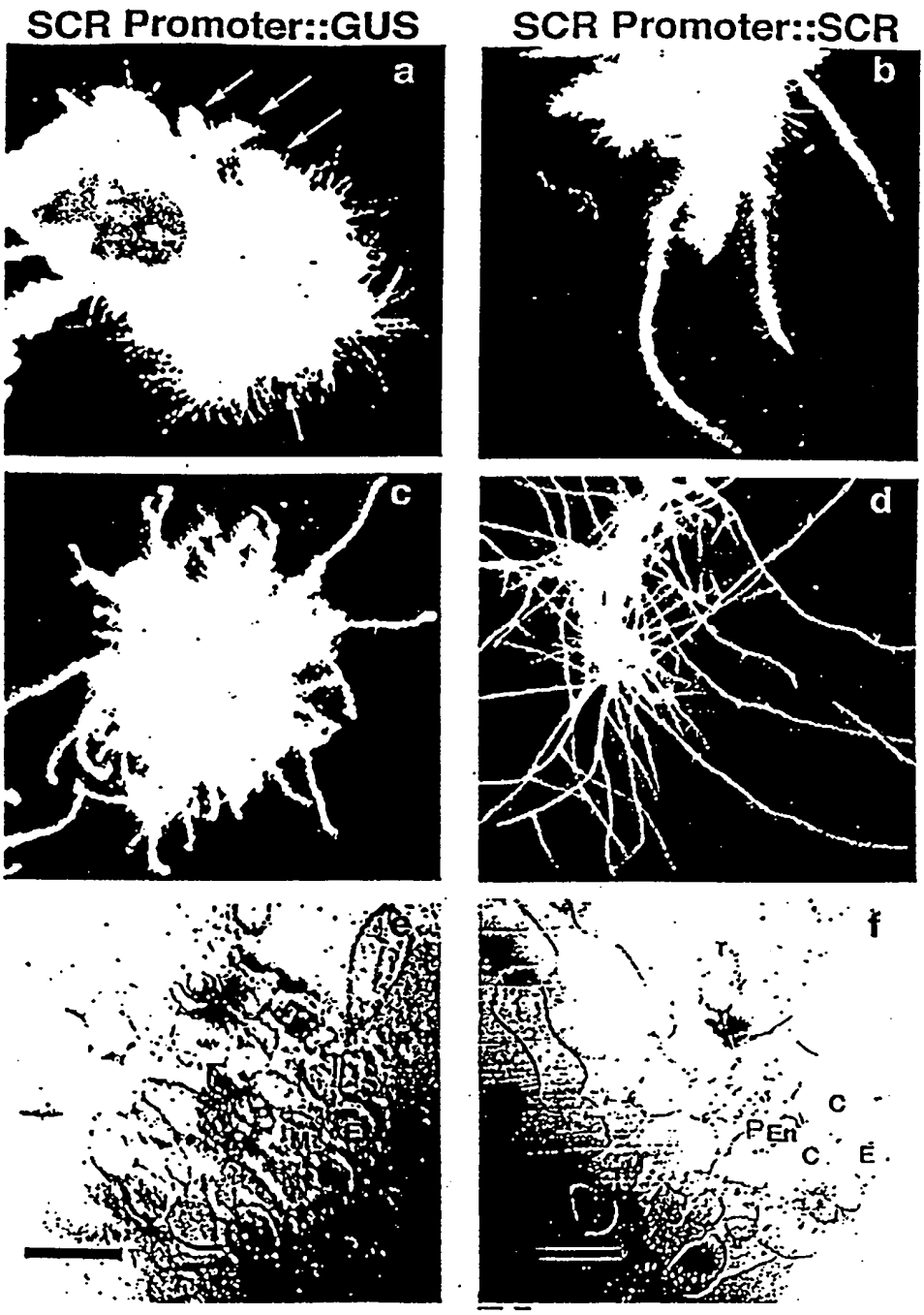


FIG. 21



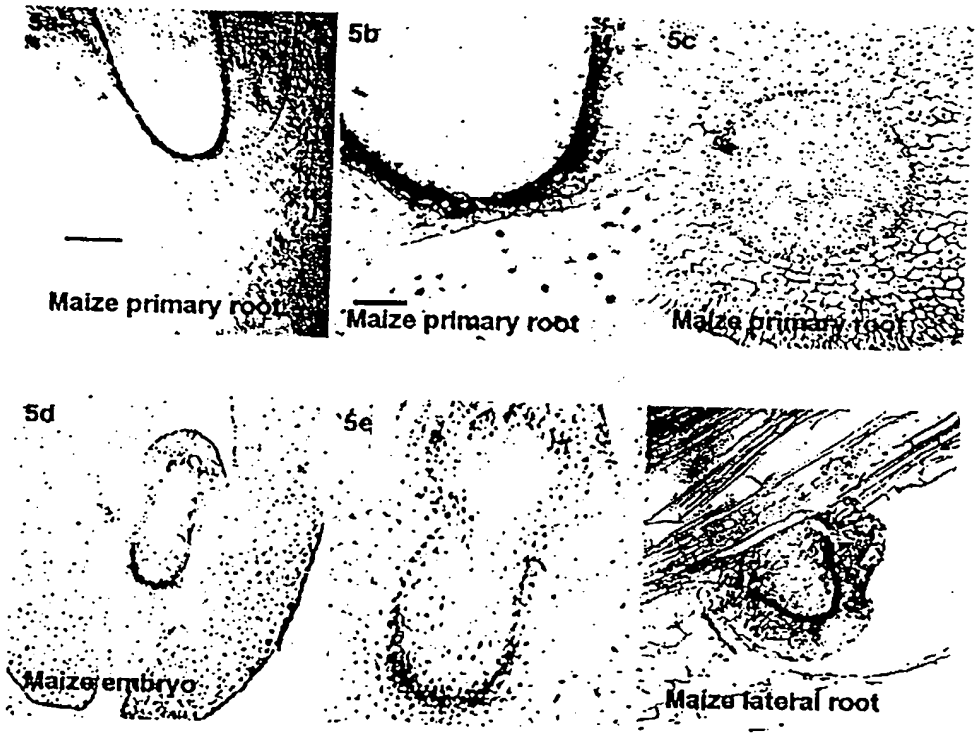
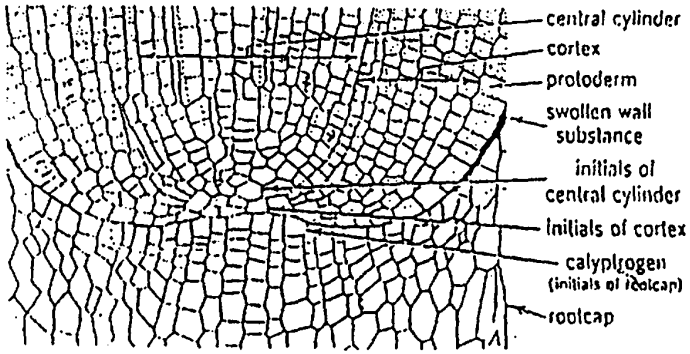
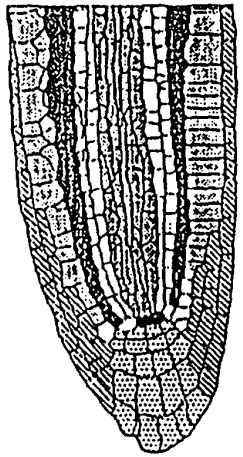

5a
5b
5c
Maize primary root
Maize primary root
Maize primary root
5d
5e
Maize embryo
Maize lateral root

Fig. 22



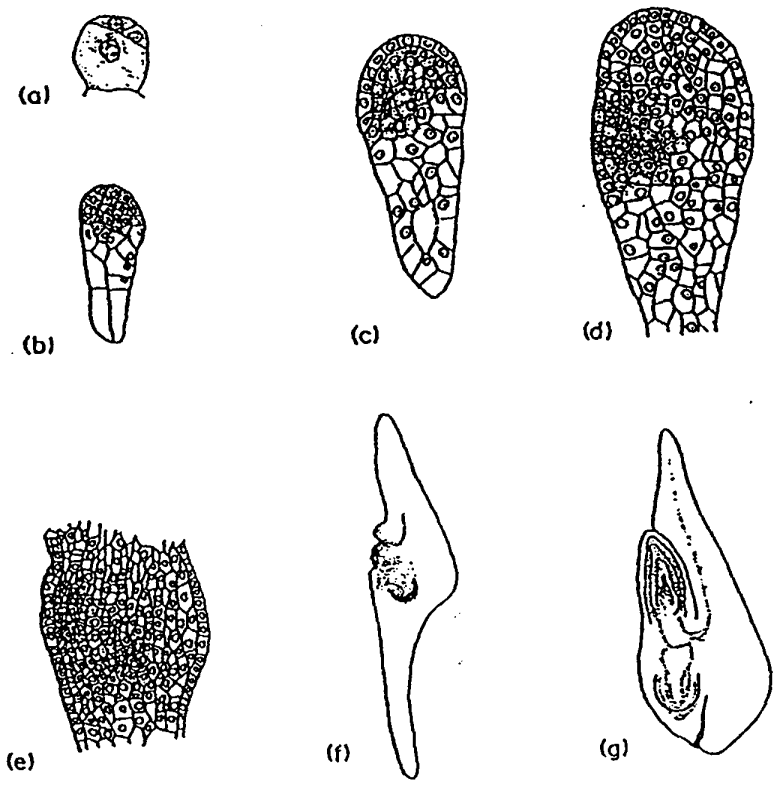
A. monocotyledonous closed-type (Zea).



B. dicotyledonous closed-type (Arabidopsis).

The schematic representation of the root apical meristems of maize (A..) And Arabidopsis (B.). Both show a type of a closed meristem in which all files of cells coverage onto a pole at the root apex, making the boundary between the root proper and the root cap discrete.

Fig. 23



Embryo development in maize. (a) Three-celled embryo showing first division of terminal cell. (b)-(c) Embryos showing embryo proper and suspensor. (d)-(e) Embryos showing the initial development of shoot and root apical meristems. (f)-(g) Embryos showing the elaborate organization of shoot and root apical meristems.

Fig. 24

```
      ctgctagctcagcctactcactccactcaactcacccccaactccactccgctcccgagc   60
      ccggactgactgactgactgtggtggtggtggtgcatcagcagcccgcgcggcgccaaaa  120
      cacgcaaactgctcctcctcactcacccctatcccccgcgctgggtcgcccgatcgcc    180
      atgcgcgcggcggcttcctcttggcgtttctagatgggctcctcctcctcctcctcttc   240
      tcctcgtcctcctccgccgcatccaccgccccccactcctttccccactctcATGCCACC  300
  1                                                       M  P  P
      GCCACCGCCTCCGCCTCCTCTCACTCCTTATTGCCGCCGCTGCCCTCCCCCACACCTCCC  360
  4    P  P  P  P  P  P  L  T  P  Y  C  R  R  C  P  P  P  H  L  P
      TCCGCCTCCTCCTTCTTCCCCAAACCACTTCCTCCTCCACTACCTCCATCAGCTAGACCA  420
 24    P  P  P  P  S  S  P  N  H  F  L  L  H  Y  L  H  Q  L  D  H
      CCAAGAAGCCGCCGCCGCCGCCATGGTCCGCAAGCGCCCCGCGTCCGACATGGACCTCCC  480
 44    Q  E  A  A  A  A  A  A  M  V  R  K  R  P  A  S  D  M  D  L  P
      GCCGCCGCGCCGCCACGTCACGGGCGACCTCTCCGACGTCACGGCGGCCGCTGCCGCCGG  540
 64    P  P  R  R  H  V  T  G  D  L  S  D  V  T  A  A  A  A  A  G
      TGTTGGTGGTAGTGGCGCGCCGTCCTCCGCCAGCGCGCAGCTGCCCGCGCTGCCCACCCA  600
 84    V  G  G  S  G  A  P  S  S  A  S  A  Q  L  P  A  L  P  T  Q
      GCTCCACCAGCTGCCCCCCGCGTTCCAGCACCACGCGCCGGAGGTGGACGTGCCCGCGCA  660
104    L  H  Q  L  P  P  A  F  Q  H  H  A  P  E  V  D  V  P  A  H
      CCCGGCCCCGGCCGCCCACGCGCAGGCGGGCGGCGAGGCAACCGCGTCCACGACCGCGTG  720
124    P  A  P  A  A  H  A  Q  A  G  G  E  A  T  A  S  T  T  A  W
      GGTGGACGGCATCATCCGCGACATCATCGGGAGCAGCGGCGGCGCCGCGGTCTCCATCAC  780
144    V  D  G  I  I  R  D  I  I  G  S  S  G  G  A  A  V  S  I  T
      GCAGCTCATCCACAACGTCCGCGAGATCATCCACCCCTGCAACCCCGGCCTCGCGTCGCT  840
164    Q  L  I  H  N  V  R  E  I  I  H  P  C  N  P  G  L  A  S  L
      CCTGGAGCTCCGCCTCCGCTCCCTCCTCGCAGCCGACCCGGCCCCACTGCCGCCGCCGCC  900
184    L  E  L  R  L  R  S  L  L  A  A  D  P  A  P  L  P  P  P  P
      GCAGCCGCAGCAGCATGCTCTCCTGCACGGCGCTCCGGCCGCCGCTCCCGCGGGGCTGAC  960
204    Q  P  Q  Q  H  A  L  L  H  G  A  P  A  A  A  P  A  G  L  T
      GCTCCCTCCCCCGCCACCGCTTCCGGACAAGCGCCGCCACGAGCATCCACCGCCGTGCCA 1020
224    L  P  P  P  P  P  L  P  D  K  R  R  H  E  H  P  P  C  Q
      GCAGCAACAGCAGGAGGAACCGCATCCGGCGCCGCAGTCGCCCAAGGCCCCGACCGCGGA 1080
244    Q  Q  Q  Q  E  E  P  H  P  A  P  Q  S  P  K  A  P  T  A  E
      AGAGACCGCAGCGGCGGCCGCCGCCGCACAAGCAGCAGCTGCTGCGGCCGCCAAGGAGCG 1140
264    E  T  A  A  A  A  A  A  A  Q  A  A  A  A  A  A  A  K  E  R
      GAAGGAGGAGCAGCGGCGGAAGCAGCGCGACGAGGAGGGCCTCCACCTGCTGACGCTGCT 1200
284    K  E  E  Q  R  R  K  Q  R  D  E  E  G  L  H  L  L  T  L  L
      GCTGCAGTGCGCCGAGGCCGTGAACGCGGACAACCTGGACGACGCGCACCAGACGCTGCT 1260
304    L  Q  C  A  E  A  V  N  A  D  N  L  D  D  A  H  Q  T  L  L
      GGAGATCGCGGAGCTAGCGACGCCGTTCGGCACCTCGACGCAGCGCGTGGCCGCCTACTT 1320
324    E  I  A  E  L  A  T  P  F  G  T  S  T  Q  R  V  A  A  Y  F
      CGCGGAGGCCATGTCGGCGCGGCTCGTCAGCTCCTGCCTGGGCCTGTACGCGCCGCTGCC 1380
344    A  E  A  M  S  A  R  L  V  S  S  C  L  G  L  Y  A  P  L  P
      GCCGGGCTCCCCCCGCCGCGGCGCGCCTCCACGGCCGCGTCGCCGCCGCGTTCCAGGTGTT 1440
364    P  G  S  P  A  A  A  R  L  H  G  R  V  A  A  A  F  Q  V  F
      CAACGGCATCAGCCCCTTCGTCAAGTTCTCGCACTTCACCGCCAACCAGGCCATCCAGGA 1500
384    N  G  I  S  P  F  V  K  F  S  H  F  T  A  N  Q  A  I  Q  E
      GGCGTTCGAGCGGGAGGAGCGCGTGCACATCATCGACCTCGACATCATGCAGGGGCTGCA 1560
404    A  F  E  R  E  E  R  V  H  I  I  D  L  D  I  M  Q  G  L  Q
      GTGGCCGGGGCTCTTCCACATCCTTGCCTCCCGCCCCGGGGGCCCGCCCAGGGTGAGGCT 1620
424    W  P  G  L  F  H  I  L  A  S  R  P  G  G  P  P  R  V  R  L
      CACCGGCCTCGGGGCGTCCATGGAGGCGCTCGAGGCCACGGGGAAGCGCCTCTCCGATTT 1680
444    T  G  L  G  A  S  M  E  A  L  E  A  T  G  K  R  L  S  D  F
      CGCCGACACGCTCGGCCTGCCCTTCGAGTTCTGCGCCGTCGCCGAGAAGGCCGGCAATGT 1740
464    A  D  T  L  G  L  P  F  E  F  C  A  V  A  E  K  A  G  N  V
```

Fig. 25A

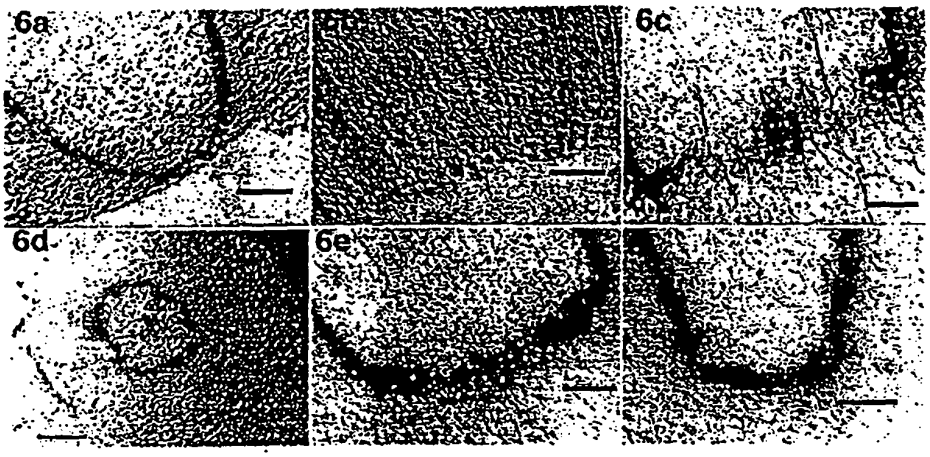
```
     TGACCCGGAGAAGCTAGGGGTCACGAGGCGGGAGGCCGTCGCCGTCCACTGGCTGCACCA  1800
484    D  P  E  K  L  G  V  T  R  R  E  A  V  A  V  H  W  L  H  H
     CTCGCTCTACGACGTCACTGGCTCCGACTCCAACACGCTCTGGCTCATCCAAAGgtagga  1860
504    S  L  Y  D  V  T  G  S  D  S  N  T  L  W  L  I  Q  R
     aggagtacaccatctctcgatcctgacttccttgctaccatgtcaaatcttgatgcaatc  1920
     atggccacttttcagctactaacactttagtttagccaatgcgacatccagtacaactaa  1980
     tctaaaaaaataatcttcagaggtttcctagtaaaaaaaccgcgtttttggagctcaaaa  2040
     agcttgtcattatgaccaaccaactttctaggcttaaaaaggttgaatcttggcaatgct  2100
     tttgagacgatgctgtactgaagtactggtagagagagtatcctccatggcctttgttga  2160
     tcccagaaccacaaaagatagtatttcgctcgcatttggttagtggaggtgttctgatca  2220
     tcacttggaggatggagctgaaagttcctatcatcatgaccaactttccatggcaaaagg  2280
     tttctagttccaagtggcaggacgatgattactgagtgactgaatggagtaactgtcatc  2340
     ttctaccactaaccatcatttattaatacataaatcatcatccggagcctaaactcagaa  2400
     aggctaatcaaaagtgcaatctttctcaaatggctgccatatgccagtggtacatgcctg  2460
     gccattgtacttttttcggtgaaccatctcgtctcaagcatgagatgaaggcctgaactgc  2520
     aatgtccttgatttgatgcaaccattattagaagaaacgctaagcgatgccggtcctggc  2580
     aagggcaatgccatatcgtcagacagacagggattcggaatcgaatggctagctggtgac  2640
     aaatcgcacggggattaataaactacattggtcattgattccatcccccacacacctgca  2700
     gGCTGGCCCCCCAAGGTGGTGACAATGGTGGAGCAGGACCTGAGCCACTCGGGCTCCTTCC  2760
522    L  A  P  K  V  V  T  M  V  E  Q  D  L  S  H  S  G  S  F
     TGGCGCGCTTCGTGGAGGCCATCCACTACTACTCGGCGCTGTTCGACTCGCTGGACGCGA  2820
541  L  A  R  F  V  E  A  I  H  Y  Y  S  A  L  F  D  S  L  D  A
     GCTACGGCGAGGACAGCCCCGAGCGGCACGTCGTGGAGCAGCAGCTGCTGTCGCGGGAGA  2880
561  S  Y  G  E  D  S  P  E  R  H  V  V  E  Q  Q  L  L  S  R  E
     TCCGCAACGTGCTGGCCGTGGGCGGGCCGGCCCGCACCGGCGACGTCAAGTTCGGCAGCT  2940
581  I  R  N  V  L  A  V  G  G  P  A  R  T  G  D  V  K  F  G  S
     GGCGCGAGAAGCTGGCGCAGTCCGGGTTCCGCGCCGCCTCGCTCGCCGGCAGCGCCGCGG  3000
601  W  R  E  K  L  A  Q  S  G  F  R  A  A  S  L  A  G  S  A  A
     CGCAGGCGTCCCTGCTGCTCGGCATGTTCCCCTCCGACGGGTACACGCTGGTGGAGGAGA  3060
621  A  Q  A  S  L  L  L  G  M  F  P  S  D  G  Y  T  L  V  E  E
     ACGGCGCGCTGAAGCTCGGGTGGAAGGACCTCTGCCTGCTCACCGCGTCGGCCTGGCGCC  3120
641  N  G  A  L  K  L  G  W  K  D  L  C  L  L  T  A  S  A  W  R
     CCATCCAGGTGCCGCCGTGCCGTTGAtgagacctctgcctgctcctgcttgcgttgagag  3180
661  P  I  Q  V  P  P  C  R  *
     gccgccactccacttgttttgcatctgtagctgctcggtttggtcatcagctgggagata  3240
     agaaaagcggaaacgtactaattgctctggagtagatccatccattcacagtgatagtta  3300
     ctgatgtactaagctttaattagttcaatgctagatcgttcttgttcaggtgtcgatcgc  3360
     gtatccttgtccttggtctcctttttcattttggtgctttgtctagtcgctttcccgacta  3420
     atgccgtgctcttcatgcgcgttctagtgaagattcttgccgagaatattagcatagttt  3480
     tcatgtaaagtagccatcaagcaagtatta  3510
```

Nucleotide and deduced amino acid sequence of the maize *SCARECROW*. Amino acid numbers are shown to the left; nucleotides are numbered on the right. Forward and reverse primers tested are underlined (J1050F and J1450R).

Fig. 25B

```
                *                  *** *      **
Zm SCR   MPPPPPPPPL TPYCRRCPPP HLPPPPPSSP NHFLLHYLHQ LDHQEAAAAA   50
At SCR   MAES------ GDFNGGQPPP HSPLRTTSSG SSSSNN--RG PPPPPPPPLV   42

         ***** ** *    *           *          *      **  * *
Zm SCR   MVRKRPASDM DLPP---PRR HVTGDLSDVT AAAAAGVGGS GAPS-SASAQ   96
At SCR   MVRKRLASEM SSNPDYNNSS RPPRRVSHLL DSNYNTVTPQ QPPSLTAAAT   92

            *                                  *       ** * *
Zm SCR   LPALPTQLHQ LP--PAFQHH APEVDVPAHP APAAH-AQAG GEATASTTAW   143
At SCR   VSSQPNPPLS VCGFSGLPVF PSDRGGRNVM MSVQPMDQDS SSSSASPTVW   142

         ** **** *  **    ***  *** *** **  **** *  * ** *****
Zm SCR   VDGIIRDIIG SSGGAAVSIT QLIHNVREII HPCNPGLASL LELRLRSLLA   193
At SCR   VDAIIRDLIH SS--TSVSIP QLIQNVRDII FPCNPNLGAL LEYRLRSLML   190

          **      *  **    *              *             * **
Zm SCR   ADPAPLPPPP QPQQHALLHG APAAAPAGLT LPPPPPLPDK RRHEHPPPCQ   243
At SCR   LDPSS-SSDP SPQTFEPLYQ ISNNPSP--- -PQQQQQHQQ QQQQHKPPPP   235

           ** *          *  *    *  * *    * *  ** *** *  *
Zm SCR   QQQQEEPHPA PQSPKAPTAE ETAAAAAAAQ AAAAAAAKER KEEQRRKQRD   293
At SCR   PIQQQERENS STDA-PPQPE TVTATVPAVQ TNTAEALRER KEEIKRQKQD   284

         ********** ******* ** **  *   ** **  * ** * ** *******
Zm SCR   EEGLHLLLTLL LQCAEAVNAD NLDDAHQTLL EIAELATPFG TSTQRVAAYF   343
At SCR   EEGLHLLLTLL LQCAEAVSAD NLEEANKLLL EISQLSTPYG TSAQRVAAYF   334

          *******   **** ** **               ***** ***** ****
Zm SCR   AEAMSARLVS SCLGLYAPLP PGSPAAARLH GRVAAAFQVF NGISPFVKFS   393
At SCR   SEAMSARLLN SCLGIYAALP SRWMPQTH-S LKMVSAFQVF NGISPLVKFS   383

         ********** *** *  *** ********** ********** ****** ***
Zm SCR   HFTANQAIQE AFEREERVHI IDLDIMQGLQ WPGLFHILAS RPGGPPRVRL   443
At SCR   HFTANQAIQE AFEKEDSVHI IDLDIMQGLQ WPGLFHILAS RPGGPPHVRL   433

         **** *****  *********  * ******* *  *** **   * * * *  *
Zm SCR   TGLGASMEAL EATGKRLSDF ADTLGLPFEF CAVAERAGNV DPEKLGVTRR   493
At SCR   TGLGTSMEAL QATGKRLSDF TDKLGLPFEF CPLAEKVGNL DTERLNVRKR   483

         ******** * *********   **** **** ***** **** *** **** *
Zm SCR   EAVAVHWLHH SLYDVTGSDS NTLWLIQRLA PKVVTMVEQD LSHSGSFLAR   543
At SCR   EAVAVHWLQH SLYDVTGSDA HTLWLLQRLA PKVVTVVEQD LSHAGSFLGR   533

         ********** ***** **** * * ****** ***** **** ******** *
Zm SCR   FVEAIHYYSA LFDSLDASYG EDSPERHVVE QQLLSREIRN VLAVGGPART   593
At SCR   FVEAIHYYSA LFDSLGASYG EESEERHVVE QQLLSKEIRN VLAVGGPSRS   583

         * *** **** *  * **    **** ** **  ********** *****  **
Zm SCR   GDVKPGSWRE KLAQSGFRAA SLAGSAAAQA SLLLGMFPSD GYTLVEENGA   643
At SCR   GEVKFESWRE KMQQCGFKGI SLAGNAATQA TLLLGMFPSD GYTLVDDNGT   633

         ******** * ****** *
Zm SCR   LKLGWKDLCL LTASAWRPIQ VPPCR   668
At SCR   LKLGWKDLSL LTASAWTPR- ----S   653
```

Fig. 26A



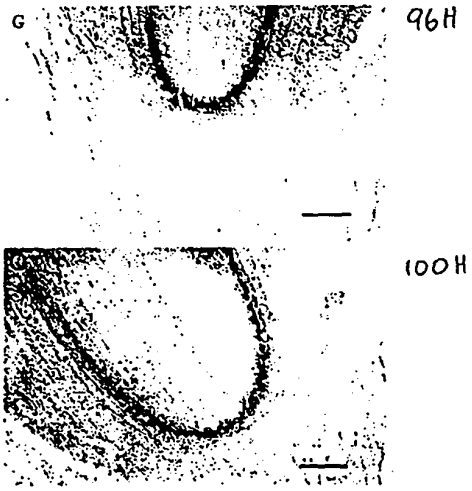


Fig. 27

Z25645

```
    ctttgtcaat ggtaaatgag ctgaggcaga tagtttctat ccaaggagac ccttctcaga
 61 gaatcgcagc ttacatggtg gaaggtctag ctgcaagaat ggccgcttca ggaaaattca
121 tctacagagc attgaaatgc aaagagcctc cttcggatga gaggcttgca gctatgagat
181 cctgtttgaa gtctgccctt gtttcaagtt cgggttttta gcagctaatg gtgcgatact
241 tgaagcaatc aaaggtgaag aagaagttca cataatcgat ttcgatataa accaagggaa
301 ccaatacatg acactgatac gaagcattgc tgagttngcc tgggtaaacg acctcgcctg
361 aggttaaaca ggaattgatg accctgaatc cagtnccaac cgctccattt gggggggcct
421 aaagaa
```

Fig. 28A

Z34183

```
    gagtacgatc ttaaagctat tcccggtgac gcgattctca atcagttcgc tatcgattcg
 61 gcttcttcgt ctaaccaagg cggcggagga gatacgtata ctacaaacaa gcggttgaaa
121 tgctcaaacg gcgtcgtgga aaccactaca gcgacggctg agatcaactc ggcatgttgt
181 cctggttgac tcgcaggaga acggtgtgcg tctcgttcac gcgcttttgg cttgcgctga
241 aagctgttca gaaagagaat ctgactgtag cggantctgg tgaagcaaat cggattctta
301 gccgtttctc aaatcggagc gatgagaaaa gtcgctactt act
```

Fig. 28B

Z34599

```
    aaatttttca attacctaat ataatgaaag ataagatctt aacaagtgac aaagggaaaa
 61 acagtaggat ttagtttggc ttcggtcgga aatctatcat cataaccggt tcaacagatc
121 aattcattga gccaccatct aattggtgag agtttccaag ccgaggtggc tatgagcggt
181 cgtgtgtgcc aacccaacat gagacagccg tcactctcct ccacccgata accctcaccg
241 ccgttgaaca gagccaaaag catactcgct tgcttaaacg cattcgaacc aatatgtgca
301 gccgcaaacc cagcagaccc gaaccggttc ctccantgac ttcaacgttt catgacggtt
361 caacttcggt ca
```

Fig. 28C

Z33772

```
    ttttttttta agtgagaacc ttaacaaatt taaccatttg aactgaaata tgaacatgta
 61 aagactcatt cacacttagc aaataggttt agaaccaaaa ctctaattat ttttatataa
121 tagggaaaaa aaagaaagaa aaattcttcc ataagtgtta gattagcttt tagtacctgt
181 gatcacccct aacctctggt aataatacat ggagatgatt taaccagtta cacaataacc
241 caagattaca gtaaaaacat aattatgttt tatgaaacat aaagactata tgctcttgtc
301 acttatctta cctccaagct gaagcaacgg attaagcttt tctcctccca gcaaaaatgg
361 gagctcaccc atttcttctt taaggttgta cttnttgca
```

Fig. 28D

Z37192

```
    gctatggaag gagagaagat ggttcatgtg attgatctcg atgcttctga gccagctcaa
 61 tggcttgctt tgcttcaagc ttttaactct aggcctgaag gtccacctca tttgagaatc
121 actggtgttc atcaccagaa ggaagtgctt gaacaaatgg ctcatagact cattgaggaa
181 gcagagaaac tcgatatccc gtttcagttt aatcccgttg tgagtaggtt agactgttta
241 aatgtagnac agtttagggt ttaaacagga gaggcnttag ccgttagctc ggttcttcaa
301 ttgcata
```

Fig. 28E

Z37191

```
    ccgatcatca aattagttat cttcagctca aattggattt ggtttggtat tacacccaca
 61 ccagaccaaa ttgaaccaac acacaaaggc tttacatgca gaggcagtag aagcatttaa
121 gccaaaatag cataaagaga cagaaagtca ccatcacaaa acaactaaga ttgtgtcccc
181 atgtatacaa aaaagaaagg gactctgctc ataaccaaaa tagaagacaa actgtaatat
241 atcattcact tcctgcatct ccaagctgat accgagtata gaggtcgatc ttgccagcaa
301 attactgcgc acccgntctc ttccttgatt ctatacccat caaaa
```

Fig. 28F

Z46550

```
    gtggaattac aattacagca atttgtattc aattgttgaa tctaagcctg gcttcatctc
 61 tttggcctgg aacgatttac ctctcctcac tctttcttcc tggcgataac caaaccaaac
121 cgatccggta ttcttagttt tgttttgttt tcaatgttat ttttggttag acaaatattc
181 aattgttaat atactccgtg gtcagagtgt tttgtttttc ttttagttcg aacgttgaat
241 taattcaggg gtaggttttg aattctctga accttatgtg ttttttggta acatcatttg
301 gatttgtgaa ctaggtttaa aaactggtct tagtcttgtt gttttctcat tagataattt
361 aaactggttt gcttctttat ttttgggttg ggataaaagt gaccgg
```

Fig. 28G

Z38048

```
    gtggaattnc aattacagca atttgtattc aattgttgaa tctaagcctg gcttcatctc
 61 tttggcctgg aacgatttac ctctcctcac tctttcttcc angcgataac caaaccaaac
121 cgatgccggt attcttagtt ttgttttgtt ttcaatgtta tttttggtta gacaaatatt
181 caattgttaa tatactccgt ggtcagagtg ttttgttttn cttttagttc gaacgttgaa
241 ttaattcagg ggtaggtttt gaattctctg aacctnatgt gttttntggt aacatcattt
301 ggatttgtga actaggttta aaaactggnc ttagtcttgt tgttttctca ttaggataat
361 ttaaactggt ttgcttcttt attttnggtt gggataaagt gaccgg
```

Fig. 28H

Z38085

```
    caaaactaca tttcatcact tttttgagca aaattacaaa taaaagagta gttacaaata
 61 tatttggctt tcaacttcct aattttatga aatagtaatt acatctcaaa cagatgacca
121 gaaccggtca ctttatccaa ccaaaaataa agaagcaaac cagtttaaat tatctaatga
181 gaaaacaaca agactaagac cagtttttaa acctagttca caaatccaaa tgatgttacc
241 aaaaaacaca taaggttcag agaattcaaa acctacccct ganttaattc aacgttcgaa
301 ctaaaagaaa aacaaaacac tctgaccacg gagtatatta acatttgatt atttgtctaa
361 ccaaaaataa cattgaaaac aaaacaaaac tanggaatac cggatcggt
```

Fig. 28I

F13896

```
    cccaacgggt cctgagcttc ttacttatat gcatatcttg tatgaagcct gcccttattt
 61 caaattcggt tatgaatctg ctaatggagc tatagctgaa gctgtgaaga acgaaagttt
121 tgtgcacatt atcgatttcc agatttctca aggtggtcaa tgggtgagtt tgatccgtgc
181 tcttggtgct agacctggtg gacctccgaa cgttaggata acgggaattg atgatccgag
241 atcatcgttt gctcgtcaag gaggacttgc agttagttgc acaaagcact tggca
```

Fig. 28J

F13897

```
    gggtcatcaa catatcactt actactacaa catttgacaa cttgttcctn cggatcatgc
 61 atgagtttta cttttacaaa cagattctgc aaactttaaa agcaagtttc taatctcttc
121 tgaaaccgaa caaggttttt attagttacc tccaagcaca agaagtgata agaggttgat
181 tcttccatcc taaatacaat gctccatctc tttcttcaag tgtatacttc tctgaataac
241 tctcaagcaa tcctttgatt gttgcgttca catacgagct caaaggatac ggtttaaatc
301 ccgccatgtg aaaccgaga
```

Fig. 28K

F13949

```
    caaaaattta tatatttgtg tgaacttaaa tttaaaaatc catcgcactg agcaaaataa
 61 nntcagaaac taaaaatttg tcatttaaga taaattgaat taaggaaaat atttttttaa
121 taattgaaac tccggtggaa atcaggagga gcgacatctc catgctgaaa ctccgacgag
181 ttctgtcctt tgccaacata ggagaagtga gttatgtttc tcctcgacgt gaaagcctct
241 cactggcgtc cgttggntna aacactcggc ttgagactcc gtgaagttac tgtgcgtcac
301 cggtgagaaa cccatctgta gaaacatcgc ttgccacgtc atcatcggcc tttctatcgg
361 acggctacga tccaacacca gcttctctat ctccggctgt ataaggaaa
```

Fig. 28L

T22782

```
    ctattttnac aatttatttt gttattagaa gtggtagtgg agtgaaaaaa caaatcctaa
 61 gcagtcctaa ccgatccccg aagctaaaga ttctncacct tcccaaataa agcaaaacct
121 agatccgaca ttgaaggaaa aacctttag atccatctct gaaaaaaacc aaccatgaag
181 agagatcatc atcatcatca tcatcaagat aagaagacta tgatgatgaa tgaagaagnc
241 gacggtaacg gcatggatga gcttctagct gttcttggtt ataaggttag gtcatccgaa
301 atggctgatg tttgctcaga aactcgagca gcttgaagtt atgatgtcta atgttcaagn
361 aagncggtct ttntcaactt cgcnacttnn gactgttcac tntaatncgg cggnngtttt
421 caacgntggc ttgntttcna tgntnaccga ccttaat
```

Fig. 28M

T21627

```
    atgggaaagg agcatttaat ctcgactcaa ttgctctacg agctctctcc ttgtttcaaa
 61 ctcggtttcg aggccgcgaa tctcgccatt ntcgacgccg ccgataacaa cgacggtgga
121 atnatgatac cgcacgtaat cgatttcaat atcggagaag gtggacaata cgttaacctt
181 ctccntacat tatccacgcg ccggaatggt aaaagtnaga gtcagaattc tccggtggtt
241 aanatcaccc gccgtggcga acaacgttta cgggatgttt agtcggatga cgggtggnga
301 agagaggttt aaaagcccgt ncgngntttt ttttgnagcc actncngntn atccg
```

Fig. 28N

H76979

```
    actcggtatc tccgtaagtt tcaacgtggt gacgagttta cgactcggtg atctgaatcg
 61 tnaatctntc gggtgtnatc ccgacgagac tttggctgta aacttagctt tcaagcttta
121 tcgtgttccc gacgaaagcg tatncacgga gaatccaaga cgaacttctc cggcgcgtga
181 agggacttaa accgcgcgtg gttactctag tggagcaaga aatgaattcg aatacggcgc
241 cgtttttagg gagagtaagt nagtcatgcg cgtttnacgg tgcgttnctt gantcggtcg
301 agtctacggt tcctagtacg gatttccgac ccgtgccaaa atttnnggaa ggaatttgcc
361 cgnaannttn naaaccgggt g
```

Fig. 28O

N96767

```
    atnaaaagtc ttttttttt ctttgttaca taagattcct acacttttcg aaatggaaaa
 61 tcacaatgat aataatatca gaataatctc gaaaattaat aataatatgg taataataag
121 aagaaaaaaa aagagtgtgt gaagttaacg ccaagcggat gcgacagtga gtgcccgtcc
181 catccaacca aagcacacac ctccgttatc ttctttaacg gtaaagcccg ggtggactcg
241 gtttccacga ctcttcatcg actccgctat cttctcactc aatggcatta actcaaaccc
301 agccatgctc atccgcattc gccatttncc ggaacanctc gnaccgctct atacgntcga
361 ttccttcgga cggcaccgng ttttactagc ttccggncaa ttccttcctn aactttggaa
421 cggtnggatt cgttcttggg accgtaggct tggcccgctt aagaacgnac cgtacagggg
481 nntgttttnt taatttccct taaaaggggg cgntttggg ttnatttttn ana
```

Fig. 28P

T43670

caaccntttt atagtcaagc agctctcaac gctttttttt caaggtctgt naagcctcga

61 aattatcaga ntttncaatc tccgtcgccg atgattganc tcacgtcggt gaatgatatg

121 agtttntttg gnggttctgg ttcatctcag cnttacggtt taccggttcc caggtctcan

181 acgcaacagc aacaatcgga ttacggttta tttggtggga tccgaatggg aatcgggtcg

241 ggtattaata attatccaac attaaccggc gttccgtgta ttgaaccggt tcaaaaccgg

301 gttcatgaat cggaggacca ttgttganta agnttaagag agctttgtng aaacaanctt

361 tttangattg atnaccg

Fig. 28Q

T76186

tgcatacaac gcaccgtttt tcgtaacacg gtttcgcgaa gtctatttca tttctcctcg

61 atttttgaca tgcttgagac aattgtgcca cgagaagacg aagagaggat gttccttgag

121 atggaggtct ttgggagaga ggcactgaat gtaattgctt gcnaaggttg ggaaagagtg

181 gagaggcctg agacatacaa gcagtggcac gtacgggcta tgaggtcagg gttggtgcag

241 gttccatttg acccaagcat tatgaagaca tcgctgcata aggtccacac attctaccac

301 aaggattttg tgatcggtca aagataaccg ggtggctctt tcaaggntgg aaggggaagg

361 anctgtcatg ggtctttctt ttttggaaac cagagtccca aggttttncc ggaaaatcct

421 ccttgggnat ttnangnccc tttttttgtt tttttncccn gnnanttccc nggggnagtt

481 tccagtttna ggngngtttt tncnaaaa

Fig. 28R

T44774

```
    tgcatacaac gcaccgtttt tngtaacacg gtttcgcgaa gtctatttna tttctcctcg
 61 atttttgaca tgcttganac aattgtncca cgagaagacg aagagaggat gttccttgan
121 atggaggtct ttgggagana ggcactgaat gtaattnctt gcnaaggttg ggaaagagtg
181 gagaggcctg anacatacaa gcagtggcac gtacgggcta tgaggtcagg gttggtgcag
241 gttccatttg acccaagcat tatgaagaca tcgctgcata aggtccacac attctaccac
301 aagggttttt tgatccntcc aagataaccg gtggctcttn caaagctttg aagggaagga
361 cctttcatgg gtcttttctt ttttggaacc aggtcccaag gttttncccg gaatccccgn
421 tggaattttg nnnccccttt tgatttttt tccccggnaa ttnccc
```

Fig. 28S

T45793

```
    gagacggtag atccgncgcg ctaaagcttc ggcgaagtaa gtagccactt tnntnatagc
 61 tccggcttga nacacagcta agcatccnat ttgcttcaca agagcttccg ctagagtcaa
121 attgtnctnc tggattgctt ctgcacaagc cataagcgcg tggactaaac gaacaccgtt
181 ctcttgcgag tnaaccagga taacagaacg anttgactca gccgccgccg tcgttgtcgt
241 ggtggttgtc gtcaccgtcg ttcctatgac tccaccaatn tgggtacccg tcgaagtcga
301 tgtaaccata ggatcagggc ttcgngcatg nttttaaaac gg
```

Fig. 28T

T46205

```
    gtttgattcg ttggaaggag ttccgaatag tcaagacaaa gtcatntctg aagtttactt
 61 agggaaacag atttgtaatc nggtggcttg tnaagntcct gacagagtcg agagacacga
121 aacgttgagt caatngggaa accggtttgg ttcgtccggt ttagcgccgg cacatcttgg
181 gtctaacgcg tttaagcaag cnagtatnct tttntntgtn tttaatagtg gccaaggtta
241 tcgtgtggag gagagtaatg gatgtttgat gttgggttgg cacactnngc ccactcattt
301 accacctccg gttttggaaa c
```

Fig. 28U

N96653

```
    taaaaattga tcccaaaaag gcataaatta aaaatgacct accaaaacga tatatataag
 61 aattttaaac aagtgaacga aaataaataa aataaacaaa aggcaaaacg gttcgattca
121 gttcggttta ggtcttggtc cgaacatatg tcatcaccgg tccactgatc tcaatctcaa
181 attcactcgn ctcgactcca ccaccgtcgt atgcttcgag tcaaactcag tacgncgccg
241 tcgagagttt ccaagcggag gtggtaatga gtggacgagt gtgccaaccc ancatcaaac
301 atccattact ttcctccaca cgntaacctt ggccactatt taaacacagg caaaangcat
361 acttgtttgc ttaaaccgcg ttagnccnaa gntttgccgg gcgntaaacc cggcngaccc
421 aanccggntt tcccnatttg ctcaaacggt ttngtgnctt ttggcttttt gnatggcctt
481 taaangnncc
```

Fig. 28V

T76483

```
    aaaaaatggg aaaccatcac tcttgatgaa cttatgatca atccaggaga gacaacggtc
 61 gtcaacngca ttcatcggtt acaatacacn cctgatgaaa ctgtgtcatt agactctcca
121 agagacacgg ttctgaagct attcagagat atcaatcctg acctctttgt gtttgcagag
181 attaacggaa tgtacaactc tcctttcttc atgacgaggt tccgagaagc gcttttncat
241 tacncttcac tctttgacat gtttgacacc acaatacacg gagaggatga gtacaaaaac
301 aggtcactgt ttggagagag agttactttt gaganacgcg nttgagcgtg attttcctgc
361 nngggnttca nancgggttt tnngggcctt aaaacctnca agaaatnggn ggtttgggtt
421 tt
```

Fig. 28W

F15454

```
    aatcaatgtt ttggttatat ttcattacta gcaacccacc cacaaccaca tgacaattta
 61 caagagaaaa acaaccacca ggtttggttt gtatacatat ataacttagg ttgtgttaca
121 acttaaaaca tcattgcaca tcctaaaaat ttcagcgacc agaatgtgtt tttgattgtg
181 cctctttctt tatccacctc aagtaaccat cattcactat aacttaccca atct
```

Fig. 28X

N37425

```
    gcgaatgttg agatcttgga agcaatagct ggggaaacca gagtccacat tatcgatttt
 61 aagattgcac agggatcaca atacatgttt ttaattcagg agcttgcgaa acgccctggt
121 gggccgccgt tgctgcgtgt nacgggtgtg gatgattcan agtccaccta tgctcgtggg
181 ggaggactca gcttggtagg tgagaggctt gcaactttgg cgcagtcatg tggtgtcccg
241 tttnagtttc acgatgccat catgtctggg tgcaaggtgc agcgggaaca tctcgggttg
301 gaacctggct ttgctgttgt tgtgaacttc ccatatgtat tacaccacat gccagacgag
361 agcgtaagtt tttgaaaatc acagngacag gcttctgcat ctnatcaana gcctttcccc
421 aaactggtac tctagtaggc aagattcaac acaacacttg catcna
```

Fig. 28Y

W43803

```
    atgnaacata tagcaaaaga tcatgcaatg agtactatat ctcttaggct acactcttac
 61 acacgctatg tcacaagcat aatataacaa cattctagtg ttcaagaacc ctaactctga
121 acttaatcca ctcgtgttgg cgagagacta tcaacagaaa agccctacat aaatcccagt
181 cgcttagaac gtaaganaca acatctatga agacgaagga acccatagag atgaagcata
241 cacgattcta cctttccacc cttgaagtaa ccagttaccg ttttgatcaa catcgaagtt
301 tttatcgtac ccgttttcgg attttcaact tcagattctg catcagttcc ttctcaagcg
361 gnagctgtcc taaatccggg tcgggtcagt ctcggctggc actggttata tggctctggg
421 ctctccactc tctctggtct tcacaaggca cancattcac aatctntttt ccataaaact
481 nnttttcntn catnngncnn atnttggctt ccctnggntg gttggggnnc ncnt
```

Fig. 28Z

W43538

```
    tcaaggttct tctttgtcat cttgttgccg aatccacaaa gaggagaata aagattcgac
 61 ctttattaga tattaacgac tctggatttt tgggtttttg gagttggatc cacatgggtt
121 cttatccgga tggattccct ggatccatgg acgagttgga tttcaataag gactttgatt
181 tgcctccctc ctcaaaccaa accttaggtt tagctaatgg gttctattta gatgacttag
241 atttctcatc cttggatcct ccagaggcat atccctccca gaacaacanc aacaacatca
301 tcaacaacaa agctgtagca ggagatctgt tatcatcttc aactgaatga cgntggattc
361 tctgattctg ttttgagtat ataagccaag ttctnatggg agnnggtnat gnagagaagc
421 ctttgtatgt tcatgnngnt ttggtnatta agntgctngg aaannactcn ntnngc
```

Fig. 28AA

SCL 1

```
LSMVNELRQI VSIQGDPSQR IAAYMVEGLA ARMAASGKFI YRALKCKEPP
SDERLAAMQV LFEVCPCFKF GFLAANGAIL EAIKGEEEVH IIDFDINQGN
QYMTLIRSIA ELPGKRPRLR LTGIDDPESV QRSIGGLRII GLRLEQLAED
NGVSFKFKAM PSKTSIVSPS TLGCKPGETL IVNFAFQLHH MPDESVTTVN
QRDELLHMVK SLNPKLVTVV EQDVNTNTSP FFPRFIEAYE YYSAVFESLD
MTLPRESQER MNVERQCLAR DIVNIVACEG EERIERYEAA GKWRARMMMA
GFNPKPMSAK VTNNIQNLIK QQYCNKYKLK EEMGELHFCW EEKSLIVASA
WR*
```

Fig. 28AB

SCL 3

AMEGEKMVHV IDLDASEPAQ WLALLQAFNS RPEGPPHLRI TGVHHQKEVL

EQMAHRLIEE AEKLDIPFQF NPVVSRLDCL NVEQLRVKTG EALAVSSVLQ

LHTFLASDDD LMRKNCALRF HNNPSGVDLQ RVLMMSHGSA AEARENDMSN

NNGYSPSGDS ASSLPLPSSG RTDSFLNAIW GLSPKVMVVT EQDSDHNGST

LMERLLESLY TYAALFDCLE TKVPRTSQDR IKVEKMLFGE EIKNIISCEG

FERRERHEKL EKWSQRIDLA GFGNVPLSYY AMLQARRLLQ GCGFDGYRIK

EESGCAVICW QDRPLYSVSA WRCRK*

Fig. 28AC

SCL 5

GTSPTGPELL TYMHILYEAC PYFKFGYESA NGAIAEAVKN ESFVHIIDFQ

ISQGGQWVSL IRALGARPGG PPNVRITGID DPRSSFARQG GLELVGQRLG

KLAEMCGVPF EFHGAALFCT EVEIEKLGVR NGEALAVNFP LVLHHMPDES

VTVENHRDRL LRLVKHLSPN VVTLVEQEAN TNTAPFLPRF VETMNHYLAV

FESIDVKLAR DHKERINVEQ HCLAREVENL IACEGVEREE RHEPLGKWRS

RFHMAGFKPY PLSSYVNATI KGLLESYSEK YTLEERDGAL YLGWKNQPLI

TSCAWR*

Fig. 28AD

SCL 6

AAIFYGHHHH TPPPAKRLNP GPVGITEQLV KAAEVIESDT CLAQGILARL

NQQLSSPVGK PLERAAFYFK EALNNLLHNV SQTLNPYSLI FKIAAYKSFS

EISPVLQFAN FTSNQALLES FHGFHRLHII DFDIGYGGQW ASLMQELVLR

DNAAPLSLKI TVFASPANHD QLELGFTQDN LKHFASEINI SLDIQVLSLD

LLGSISWPNS SEKEAVAVNI SAASFSHLPL VLRFVKHLSP TIIVCSDRGC

ERTDLPFSQQ LAHSLHSHTA LFESLDAVNA NLDAMQKIER FLIQPEIEKL

VLDRSRPIER PMMTWQAMFL QMGFSPVTHS NFTESQAECL VQRTPVRGFH

VEKKHNSLLL CWQRTELVGV SAWRCRSS*

Fig. 28AE

SCL 11

KKWETITLDE LMINPGETTV VNCIHRLQYT PDETVSLDSP RDTVLKLFRD

INPDLFVFAE INGMYNSPFF MTRFREALFH YSSLFDMFDT TIHCERRDEV

ISCEGAERFA RPETYKQWRV RILRAGFKPA TISKQIMKEA KEIVRKRYHR

DFVIDSDNNW MLQGWKGRVI YAFSCWKPAE KFTNNNLNI*

Fig. 28AF

SCL 13

ANVEILEAIA GETRVHIIDF QIAQGSQYMF LIQELAKRPG GPPLLRVTGV

DDSQSTYARG GGLSLVGERL ATLAQSCGVP FEFHDAIMSG CKVQREHLGL

EPGFAVVVNF PYVLHHMPDE SVSVEKYRDR LLHLIKSLSP KLVTLVEQES

NTNTSPLVSR FVETLDYYTA MFESIDAARP RDDKQRISAE QHCVARDIVN

MIACEESERV ERHEVLGKWR VRMMMAGFTG WPVSTSAAFA ASEMLKAYDK

NYKLGGHEGA LYLFWKRRPM ATCSVWKPNP NYIG*

Fig. 28AG

SCL 14

LLKVLLCHLV AESTKRRIKI RPLLDINDSG FLGFWSWIHM GSYPDGFPGS

MDELDFNKDF DLPPSSNQTL GLANGFYLDD LDFSSLDPPE AYPSQNNNNN

NINNKAVAGD LLSSSSDDAD FSDSVLKYIS QVLMEEDMEE KPCMFHDALA

LQAAEKSLYE ALGEKDPSSS SASSVDHPER LASHSPDGSC SGGAFSDYAS

TTTTTSSDSH WSVDGLENRP SWLHTPMPSN FVFQSTSRSN SVTGGGGGGN

SAVYGSGFGD DLVSNMFKDD ELAMQFKKGV EEASKFLPKS SQLFIDVDSY

IPMNSGSKEN GSEVFVKTEK KDETEHHHHH SYAPPPNRLT GKKSHWRDED

EDFVEERSNK QSAVYVEESE LSEMFDNMFL CGPGKPVCIL NQNFPTESAK

VVTAQSNGAK IRGKKSTSTS HSNDSKKETA DLRTLLVLCA QAVSVDDRRT

ANVXLRQIRE HSSPLGNGSE RLAHYFANSL EARLAGTGTQ IYTALSSKKT

SAADMLKAYQ TYMSVCPFKK AAIIFANHSM MRFTANANTI HIIDFGISYG

FQWPALIHRL SLSRPGGSPK LRITGIELPQ RGFRPAEEFR RQVIAWLDTV

SDTMFRLSTT QLLRNGETIQ VEDLKLRQGE YVVVNSLFRF RNLLDETVLV

NSPRDAVLKL IRKINPNVFI PAILSGNYNA PFFVTRFREA LFHYSAVFDM

CDSKLAREDE MRLMYVFEFY GREIVNVVAS EGTERVESRE TYKQWQARLI

RAGFRQLPLE KELMQNLKLK IENGYDKNFD VDQNGNWLLQ GWKGRIVYAS

SLWVPSSS*

Fig. 28AH

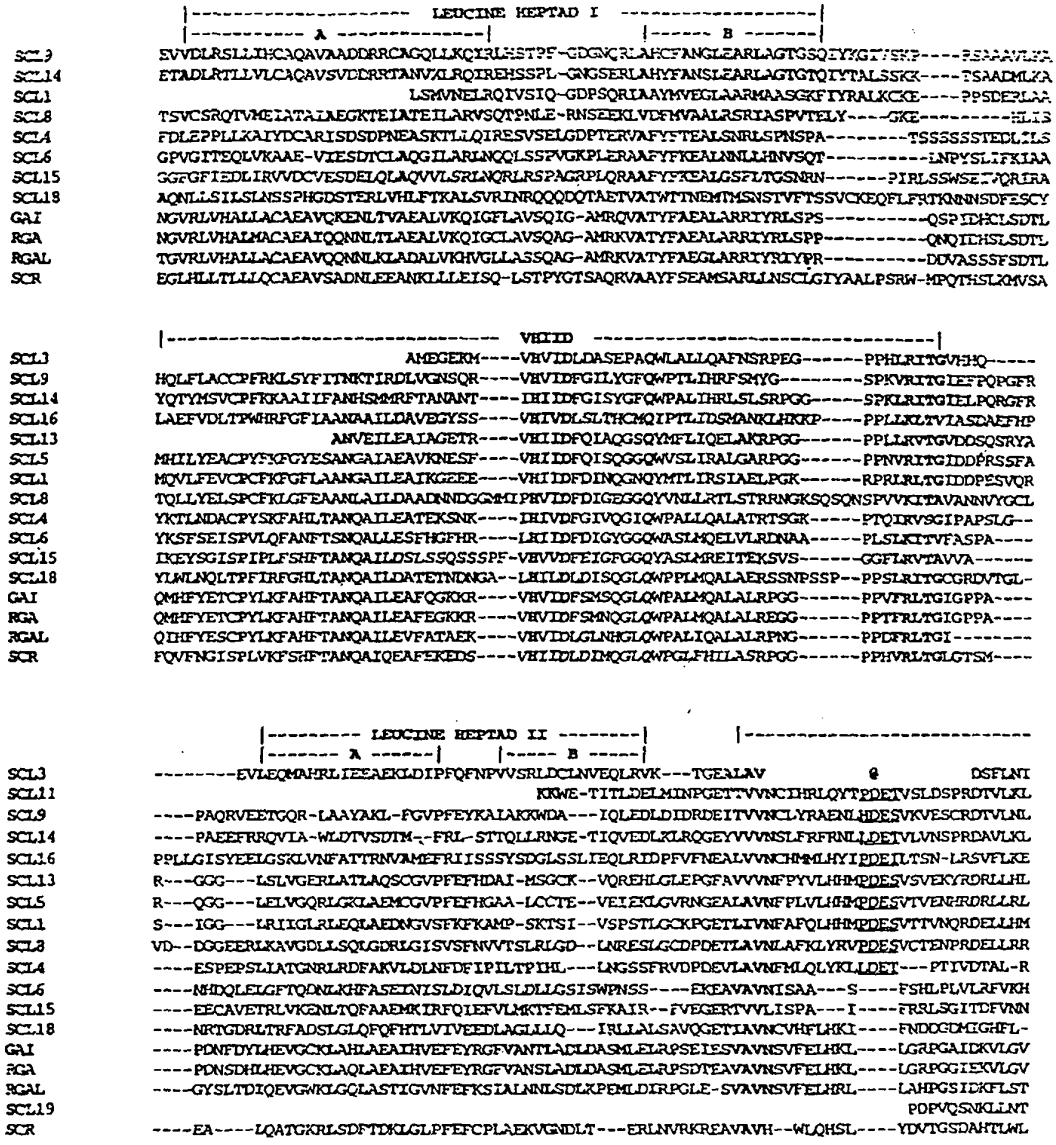

LEUCINE HEPTAD I
A
B
VHIID
LEUCINE HEPTAD II
A
B
SCL3
SCL9
SCL14
SCL1
SCL8
SCL4
SCL6
SCL15
SCL18
GAI
RGA
RGAL
SCR
SCL16
SCL13
SCL5
SCL11
SCL12
SCL19

Fig. 29A

@SSVLQLHTFLASDDDLMRKNCALRFHNNPSGVDLQRVLMMSHGSAAEARENDMSNNNGYSPSGDSASSLPLPSSGRT

Fig. 29B

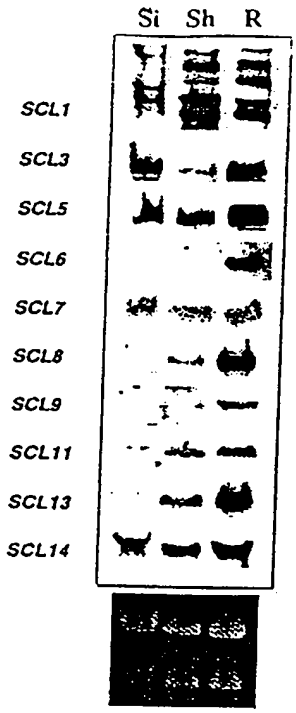


Fig. 30

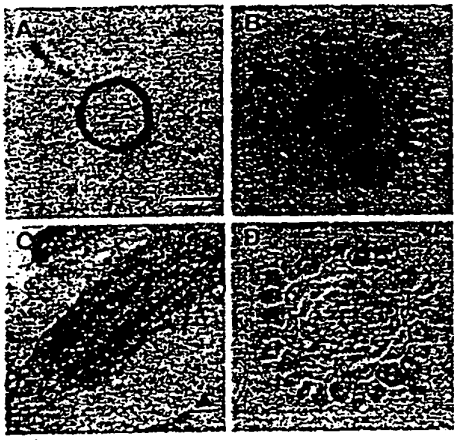
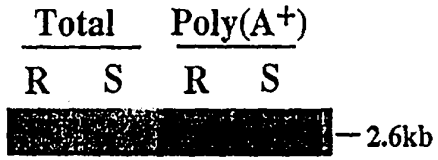


Fig. 31

RNA Blot Analysis



Either total RNA or poly (A^+) RNA was probed with the full length of cDNA.
About 2.6kb fragment was hybridized to the probe.
R: Roots, S: Shoots

Fig. 32

CBPBT44 partial cDNA sequence

GCGGCCGCGCAGAGCCGCCGCGTGGCGGTGGCGTTCCAGGCGTACAACGCGCTGTCGCCG
CTCGTCAAGTTCTCGCACTTCACGGCCAACCAGGCCATCCTGCAGGCGCTCGACGGCGAG
GACTGCCTCCACGTGATCGACCTGGACATCATGCAGGGCCTGCAGTGGCCGGGGCTCTTC
CACATCCTCGCGTCCCGCCCGCGCAAGCCGCGGTCGCTCCGGATCACCGGGCTCGGCGCG
TCGCTCGACGTCCTCGAGGCCACTGGCCGCCGCCTCGCCGACTTCGCGGCCTCGCTCGGC
CTCCCGTTCGAGTTCCGCCCCATCGAGGGGAAGATCGGGCACGTCGCCGACGCCGCGGCG
CTCCTCGGCTCGCGCCAGCGGCGGCGGGATGACGAGGCCACCGTGGTGCACTGGATGCAC
CACTGCCTCTATGACGTGACGGGGTCGGACGTGGGCACGGTGCGGCTGCTCCGGAGCCTG
CGCCCGAAGCTGATCACCATCGTGGAGCAGGACCTGGGCCACAGCGGGGATTTCCTGGGC
CGGTTCGTGGAGGCGCTGCACTACTACTCGGCGCTGTTCGACGCGCTGGGAGACGGCGCC
GGCGCGGCCGAGGAGGAGTCGGCCGAGCGGTACGCGGTTGAGCGACAGCTCCTGGGCGCG
GAGATACGCAACATCGTGGCCGTAGGGGGGCCCAAGCGGACAGGGGAGGTGCGCGTGGAG
CGGTGGAGCCACGAACTGCGGCACGCCGGGTTCCGGCCAGTGTCCCTGGCCGGGAGCCCT
GCCGCGCAGGCCAGGCTGCTCCTCGGCATGTATCCGTGGAAGGGGTACACGCTGGTGGAG
GAGGACGCGTGCCTTAAGCTGGGCTGGAAGGACCTCTCCCTGCTCACCGCGTCGGCGTGG
GAGCCGGCGGACGACGCTGCCGCTTCTGCGCCCACCGGTTAACGAGTACGAGCGGACGCG
TGGGTCGAC

CBPBT44 partial amino acid sequence

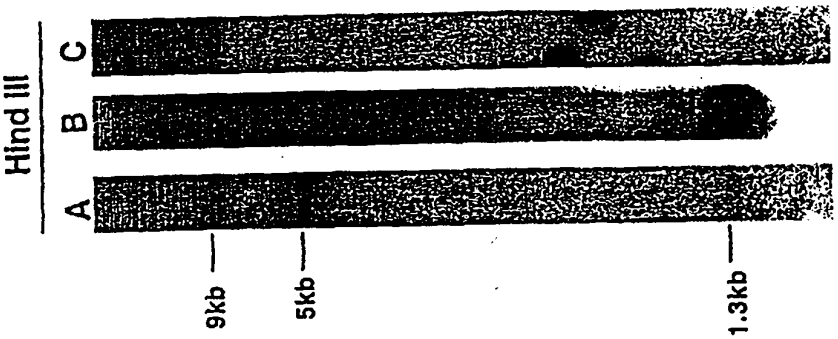
AAAQSRRVAVAFQAYNALSPLVKFSHFTANQAILQALDGEDCLHVIDLDIMQGLQWPGLF
HILASRPRKPRSLRITGLGASLDVLEATGRRLADFAASLGLPFEFRPIEGKIGHVADAAA
LLGSRQRRRDDEATVVHWMHHCLYDVTGSDVGTVRLLRSLRPKLITIVEQDLGHSGDFLG
RFVEALHYYSALFDALGDGAGAAEEESAERYAVERQLLGAEIRNIVAVGGPKRTGEVRVE
RWSHELRHAGFRPVSLAGSPAAQARLLLGMYPWKGYTLVEEDACLKLGWKDLSLLTASAW
EPADDAAASAPTGXRVRADAWVD

Fig. 33

```
Zm SCR                                                GRVAAAFQVF NGISPFVKFS
CBPBT44                                               RRVAVAFQAY NALSPLVKFS
At SCR                                                LKMVSAFQVF NGISPLVKFS

Zm SCR   HFTANQAIQE AFEREERVHI IDLDIMQGLQ WPGLFHILAS RPGGPPRVRL
CBPBT44  HFTANQAILQ ALDGEDCLHV IDLDIMQGLQ WPGLFHILAS RPRKPRSLRI
At SCR   HFTANQAIQE AFEKEDSVHI IDLDIMQGLQ WPGLFHILAS RPGGPPHVRL

Zm SCR   TGLGASMEAL EATGKRLSDF ADTLGLPFEF CAVAEKAGNV DPEKLGVTRR
CBPBT44  TGLGASLDVL EATGRRLADF AASLGLPFEF RPIEGKIGHV ADAAALLGSR
At SCR   TGLGTSMEAL QATGKRLSDF TDKLGLPFEF CPLAEKVGNL DTERLNVRKR

Zm SCR   ------EAVA VHWLHHSLYD VTGSDSNTLW LIQRLAPKVV TMVEQDLSHS
CBPBT44  QRRRDDEATV VHWMHHCLYD VTGSDVGTVR LLRSLRPKLI TIVEQDLGHS
At SCR   ------EAVA VHWLQHSLYD VTGSDAHTLW LLQRLAPKVV TVVEQDLSHA

Zm SCR   GSFLARFVEA IHYYSALFDS LDASYGEDSP ERHV---VEQ QLLSREIRNV
CBPBT44  GDFLGRFVEA LHYYSALFDA LGDGAGAAEE ESAERYAVER QLLGAEIRNI
At SCR   GSFLGRFVEA IHYYSALFDS LGASYGEESE ERHV---VEQ QLLSKEIRNV

Zm SCR   LAVGGPARTG DVKFGSWREK LAQSGFRAAS LAGSAAAQAS LLLGMFPSDG
CBPBT44  VAVGGPKRTG EVRVERWSHE LRHAGFRPVS LAGSPAAQAR LLLGMYPWKG
At SCR   LAVGGPSRSG EVKFESWREK MQQCGFKGIS LAGNAATQAT LLLGMFPSDG

Zm SCR   YTLVEENGAL KLGWKDLCLL TASAWRPIQV PPCR
CBPBT44  YTLVEEDACL KLGWKDLSLL TASAWEPADD AAASAPTG
At SCR   YTLVDDNGTL KLGWKDLSLL TASAWTPRS
```

Fig. 34

DNA Blot Analysis



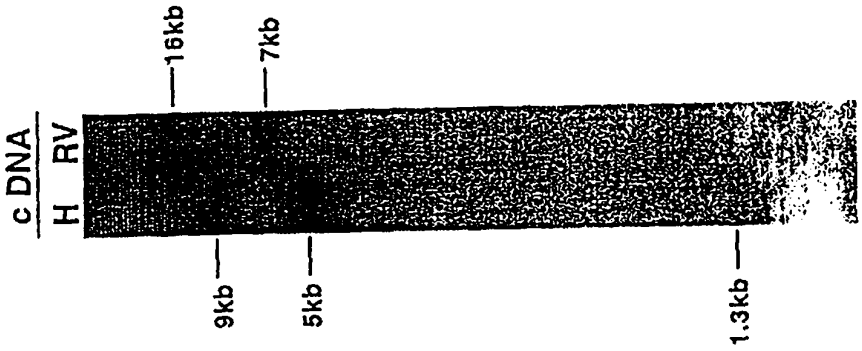


Fig. 35